Medical Teacher, Vol. 25, No. 6, 2003, pp. 569–584

Taylor&Francis healthsciences

AMEE Guide No. 25: The assessment of learning outcomes for the competent and reflective physician

J.M. SHUMWAY[1] & R.M. HARDEN[2]

[1]*Office of Medical Education, West Virginia University, USA;*
[2]*International Virtual Medical School, 484 Perth Road, Tay Park House, Dundee, UK*

SUMMARY *Two important features of contemporary medical education are recognized. The first is an emphasis on assessment as a tool to ensure quality in training programmes, to motivate students and to direct what they learn. The second is a move to outcome-based education where the learning outcomes are defined and decisions about the curriculum are based on these. These two trends are closely related. If teachers are to do a better job of assessing their students, they need an understanding of the assessment process, an appreciation of the learning outcomes to be assessed and a recognition of the most appropriate tools to assess each outcome. Assessment tools selected should be valid, reliable, practical and have an appropriate impact on student learning. The preferred assessment tool will vary with the outcome to be assessed. It is likely to be some form of written test, a performance test such as an OSCE in which the student's competence can be tested in a simulated situation, and a test of the student's behaviour over time in clinical practice, based on tutors' reports and students' portfolios. An assessment profile can be produced for each student which highlights the learning outcomes the student has achieved at the required standard and other outcomes where this is not the case. For educational as well as economic reasons, there should be collaboration across the continuum of education in test development as it relates to the assessment of learning outcomes and in the implementation of a competence-based approach to assessment.*

The importance of assessment

Assessment plays a major role in the process of medical education, in the lives of medical students, and in society by certifying competent physicians who can take care of the public. The very foundation of medical curricula is built around assessment milestones for students. For example, in the United States medical students must pass a series of steps towards licensure before graduating from medical school. It is assessment and evaluation that often drives the curricula of medical schools and students measure their progress through the curriculum by the examinations they have passed. Assessment becomes a motivating force for them to learn. Society has the right to know that physicians who graduate from medical school and subsequent residency training programmes are competent and can practise their profession in a compassionate and skilful manner. It is the responsibility of the medical school to demonstrate that such competence has been achieved, and the responsibility of accreditation agencies to certify that the educational programmes in medical schools can do what they promise. Assessment is of fundamental importance because it is central to public accountability.

The General Medical Council (GMC) has the responsibility to ensure that graduates of a UK medical school have met the requirements for their next posts as house officers. In 1993 they issued their recommendations on undergraduate medical education (GMC, 1993). More recent recommendations (GMC, 2002) place greater emphasis on learning outcomes and on the assessment of the outcomes. ''In line with current educational theory and research, we have adopted an outcomes-based model. This sets out what is to be achieved and assessed at the end of the medical course in terms of knowledge, skills and behaviour'' (Rubin & Franchi-Christopher, 2002). Table 1 contains a summary of those recommendations as they relate to assessment.

In the United States, the Liaison Committee for Medical Education (LCME) is the accreditation agency for North American medical schools (USA and Canada). Medical schools in North America have traditionally been accredited on the quality of the elements that make up the student educational programme (e.g. faculty, research, facilities, courses and clerkships). There are essentially four questions asked during accreditation: (1) What are the goals?; (2) What did students actually learn?; (3) What is the evidence?; and (4) What needs to be changed? The LCME has instituted standards focusing on the assessment of outcomes (LCME, 2003). In outcome-based assessment the educational programme goals or learning outcomes are defined and their accomplishment is assessed. North American medical education institutions are now required to document educational outcomes in light of their institutional purposes and missions. The LCME standards pertaining to the assessment of these outcomes are included in Table 2.

Outcome-based assessment for a competent and reflective physician

Assessment is an intrinsic component of outcome-based education. Outcome-based education and performance assessment are closely related paradigms (Friedman Ben-David, 1999). Outcome-based education involves an educational approach in which the decisions about the curriculum and evaluation are driven by the learning outcomes that students should achieve (Harden *et al.*, 1999 a). In this approach, the product (student learning outcomes) defines the process (instructional methods and learning opportunities). This is distinctively different from earlier educational

Correspondence: Professor James M. Shumway, Office of Medical Education, West Virginia University, School of Medicine, PO Box 9007, Morgantown, WV 26506, USA. Email: jshumway@hsc.wvu.edu

ISSN 0142–159X print/ISSN 1466–187X online/03/060569-16 © 2003 Taylor & Francis Ltd
DOI: 10.1080/0142159032000151907

Table 1. Recommendations of the General Medical Council in the UK relating to assessing student performance and competence (GMC, 2002).

The Principles of Assessment

Schemes of assessment must support the curriculum and allow students to prove that they have achieved the curricular outcomes. This means assessment must allow students to demonstrate the breadth and depth of their knowledge, and to show what they can do. Professional attitudes and behaviour must also be assessed.

Medical schools should use a range of assessment techniques that are appropriate for testing the curricular outcomes. Medical schools should determine the most appropriate scheme of assessment for their curriculum. However, schemes must meet best practice in assessment, and medical schools must be able to provide evidence that the schemes are valid and reliable, and that they have processes for setting standards and making decisions about student performance.

When students get close to graduating, their knowledge, skills, attitudes and behaviour must be thoroughly assessed to determine their fitness to practice as pre-registration house officers.

Assessment Procedures

Schemes of assessment must be open, fair and meet appropriate standards. Medical schools must make sure that:

a. there is a clear indication of how the scheme of assessment deals with all the curricular outcomes;
b. there is a clear indication of how individual assessments and examinations contribute to the overall assessment of the curricular outcomes;
c. when they design individual assessments, there is a clear indication of how the targeted curricular outcomes have been met;
d. students have clear guidance about what is expected of them in any examination or assessment;
e. examiners are trained to carry out the role and to apply the medical school's assessment criteria consistently;
f. examiners have clear guidelines for marking assessments, which indicate how performance against targeted curricular outcomes should be rewarded;
g. systems are in place to determine the pass mark; and
h. external examiners are employed to make sure that standards are met.

Note: The methods are listed for each of the assessment outcomes in order of importance.

Table 2. LCME standards related to the assessment of learning outcomes (LCME, 2003).

''Medical schools must evaluate educational program effectiveness by documenting the achievements of their students and graduates in verifiable and internally consistent ways that show the extent to which institutional and program purposes are met.''

''Medical schools should use a variety of measures to evaluate program quality, such as data on student performance/achievement, acceptance into residency programs, postgraduate performance and licensing, and emerging measures that may prove to be valid.''

''The results of such evaluations should be used to determine how well schools are fulfilling their objectives and to assess the need for program improvement.''

''Schools also should evaluate the performance of their students and graduates in the framework of national norms of accomplishment.''

approaches that relied on inputs. These approaches, having defined the educational programme, accepted whatever outcomes resulted from the process. The assumption was that a 'better' process would result in 'better' outcomes. In outcome-based education, agreement on predetermined student learning outcomes defines the processes used to achieve them. Put simply, outcome-based education has two requirements:

(1) Expected learning outcomes are initially agreed and then communicated to all involved in the educational process (students, instructors, employers, the public, etc.).
(2) Learning outcomes determine curriculum content, teaching methods and assessment.

All decisions concerning the curriculum are based on achieving the desired learning outcomes.

Outcome-based education programmes, argued Friedman Ben-David (1999), are faced with the need to develop non-traditional teaching and assessment techniques which capture both the learning and performance of broad abilities. The clearer the definition of learning outcomes, the more effectively can the assessment process be planned and implemented.

Over the past 10 years, educators and individuals in a number of countries have been developing approaches to outcome-based education (Harden, 2002). In the USA, Kassebaum *et al.* (1997) reviewed accreditation reports of medical schools and found that only a small percentage of schools had 'robust' institutional objectives that guided their educational programmes. They noted that schools where there was a lack of institutional objectives were more likely to have accreditation citations for shortcomings in curricular management. The Association of American Medical Colleges

(AAMC) formed an advisory group to recommend guidelines to US medical schools on the objectives of medical education. The Medical Schools Objectives Project (MSOP) identified attributes based on society's expectations of a good physician. They were grouped into four categories: (1) physicians must be altruistic; (2) physicians must be knowledgeable; (3) physicians must be skilful; and (4) physicians must be dutiful (AAMC, 1998). Some individual medical schools in the USA have developed their own competences. Brown University in Providence, Rhode Island, described a list of nine abilities (Smith & Dollase, 1999; Smith, 1999). Likewise, in Canada, physician groups developed the essential competences and roles of their profession. The CanMEDS 2000 Project Societal Needs Working Group reported seven roles of specialist physicians: (1) medical expert, (2) communicator, (3) collaborator, (4) manager, (5) health advocate, (6) scholar, and (7) professional (CanMEDS, 2000). Within each of these categories, a number of specific attributes or objectives were identified. The Accreditation Council for Graduate Medical Education (ACGME) in the USA also specified learning outcomes for the training of the doctor (ACGME, 2003). The International Institute of Medical Education has produced an international consensus on the learning outcomes or minimum essential requirements expected of a student on graduation from medical school (IIME, 2002).

In the UK, all five Scottish medical schools have adopted the same framework for learning outcomes (Simpson *et al.*, 2002). This is based on the three-circle model, devised to classify the learning outcomes at the University of Dundee School of Medicine as illustrated in Figure 1 (Harden *et al.*, 1999a, 1999b). The model is based on the three essential aspects of competence of a generalist physician. The inner sphere describes those things the physician is able to do. These include the clinical, procedural, investigation, management, health promotion, communication, and information-handling skills. The middle layer represents how the physician approaches the skills with knowledge and understanding, ethical/legal principles, and clinical reasoning and decision-making abilities. The outer layer represents the development of the personal characteristics of the physician. Such characteristics include understanding physicians' roles in society and their personal development as lifelong learners and professionals. This model has also been used in other countries such as Spain and Sweden and in the training of junior doctors and specialists in the UK.

The outcomes specified by these different bodies have similarities and embrace a similar set of competences or abilities. For the purpose of this guide we have used the 12 learning outcomes as identified in the Dundee three-circle model.

Aim of this guide

Many scholars in the field of assessment have contributed to descriptions of approaches to assessment and the assessment instruments that are referred to in this guide. Less has been written, however, on the application of these assessment approaches to the specific learning outcomes of medical education (ACGME and ABMS, 2000; Scottish Deans' Medical Curriculum Group, 2002).

A critical examination of practice relating to student assessment should be undertaken in the context of the changing expectations with regard to students' learning outcomes. There has been a change in emphasis from knowledge acquisition and factual recall to more widely embracing learning outcomes such as problem solving, clinical judgement, communication skills, attitudes and professionalism. To date, however, these changes in the curriculum have not been fully reflected in the assessment process.

In the United States, multiple-choice questions continue to play a major role as an assessment tool, especially in the early years of the curriculum. While clinical and technical skills are increasingly assessed by approaches such as the objective structured clinical examination (OSCE) and the use of standardized patients (SPs), little has been done to assess attitudes and other professional behaviours and characteristics of physicians in training. Such mismatches between the choice of assessment tools and the expected learning outcomes cannot be attributed to the lack of appropriate assessment instruments. An extensive number of assessment instruments is now available to the examiner.

The range of instruments from which examiners can make their choice includes traditional approaches such as written questions, approaches introduced over the past few decades such as the OSCE and relative newcomers to the assessment scene such as portfolios.

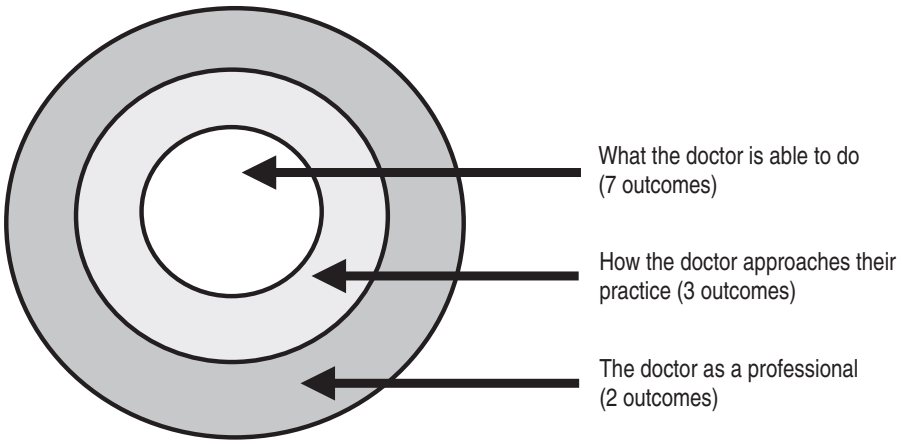

Figure 1. The three-circle framework for learning outcomes.

In designing and implementing a programme of student assessment, examiners must be knowledgeable about the expected learning outcomes and the methods available which might be applied to their measurement. They must also have an understanding of the assessment process in order to allow them to make an appropriate choice of methods and implement them in practice. It is important to understand how to use new assessment tools, and how to evaluate their strengths, weaknesses and applicability under various instructional conditions.

This guide describes assessment approaches for each of the specific competences expected of students or trainees (learners). All schools and institutions engaged in the move to competence-based approaches to medical education, with the associated need to assess expected learning outcomes, should find the suggestions helpful.

The guide takes a multi-dimensional view of assessment (Figure 2). It describes how as educators we can do a better job of assessing students by assessing the full range of learning outcomes, by choosing appropriate instruments for the purpose and by having an understanding of the assessment process.

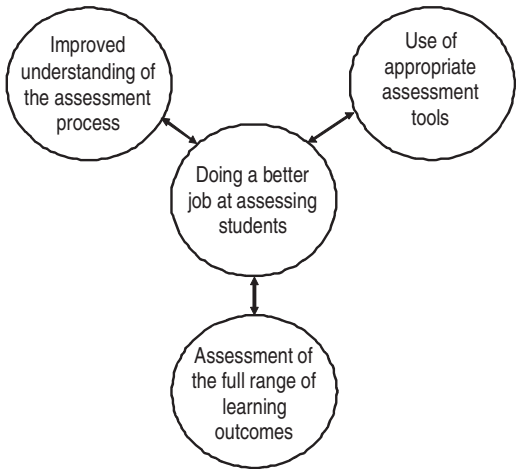

Figure 2. A multi-dimensional model of assessment.

Improved understanding of the assessment process

Assessment instruments can be described according to certain prescribed criteria that are evidence based and recognized by professionals in the field. The criteria most commonly referred to are: (1) validity, (2) reliability, (3) impact on the learner and educational programme, and (4) practicality including cost.

Validity

The validity of an instrument is the degree to which an instrument measures what it is supposed to measure (Goetz *et al.*, 1992; Atkinson *et al.*, 1993; Pressley & McCormick, 1995). It is concerned with whether there is anything about a test that affects an examinee's score so that the test fails to measure the learning outcomes intended. Unfortunately, a gold standard to measure the validity of an assessment instrument against is often unavailable. For assessment instruments, validity concerns itself with a specific measurement in a specific situation with a specific group of individuals. In other words, it is not the instrument or approach that is valid but the scores that are valid (ACGME and ABMS, 2000). What is being measured depends as much on the content of the assessment as on any characteristic of the method (Van der Vleuten, 1996).

Validity can be broken down into three major categories: content validity, criterion-related validity and construct validity. Content validity is determined by a review of the assessment instrument and the extent to which it measures what it is intended to measure. Criterion-related validity refers to a comparison of the test score against a known criterion of the expected performance. Construct validity refers to a collection of indirect information that the assessment instrument measures what it purports to measure. For example, construct validity is corroborated if the instrument is able to distinguish between different levels of trainees (Winckel *et al.*, 1994). A simple way to remember the concept of validity is to think of a 'bulls-eye target', such as used in darts and archery. The degree of accuracy to which items end up in the desired central targeted area indicates the degree of validity of measuring what was intended.

Reliability

The reliability of a test is the consistency, generalizability or reproducibility of a test. It is the extent to which examinees' scores would be similar if they were retested. Reliability is the degree to which an instrument consistently measures the phenomena over time (Goetz *et al.*, 1992; Atkinson *et al.*, 1993; Pressley & McCormick, 1995). If the test is repeated over time, then the new test results should be similar to the earlier scores for the same assessment instrument on the same group of individuals (ACGME and ABMS, 2000). This is called the stability of test scores or test–retest reliability. Another approach to estimating reliability is to prepare two forms of a test and compare the results. This is called equivalent or alternative-forms reliability. Internal consistency is the reliability approach most typically practised. It involves the extent to which a student's scores on two or more portions of a test agree with each other (Goetz *et al.*, 1992). Split-half reliability is the simplest approach for measuring internal consistency.

Reliability is usually measured as a correlation coefficient with 1.0 being a perfect correlation and zero being no correlation. Values above 0.70 are considered to indicate a reliable instrument although some aim for a figure above 0.80. Inter-rater reliability refers to the level of agreement among different independent evaluators of the same examinee (Winckel *et al.*, 1994). An instrument is unlikely to be reliable if it has only a small number of items sampling the intended behaviour. So reliability is a matter of adequately sampling the content. Reliability is also a function of the time spent in the test setting (Van der Vleuten, 2000). Again, thinking of a bulls-eye target is also helpful with regard to reliability. If the results of the assessment cluster around each other like the arrows on the target, then it can be said that the evaluations are reliable, even if the cluster is off centre of the target.

Impact on learning

An important but often ignored characteristic of an evaluation instrument is its impact on the learner. Nowhere is this

more true than in medical education. An old adage in medical education is that 'examinations drive the curriculum'. The type of examination can determine how students study, thus how they learn. It is also true that the importance and frequency of examinations also impact on the student's learning environment. Student responses to examinations may include memorization, studying at the last moment, concentrating on their perception of 'what is the right answer', and cheating. Van der Vleuten (2000) has emphasized that medical educators need to pay more attention to these characteristics than is currently done. Van der Vleuten (1996) has described at least four ways that assessment can drive the learning process. Assessment drives learning through: (1) its content, (2) its structure or format, (3) what is asked (e.g. the information given), and (4) its frequency, timing and the number of repeat examinations.

Practicality including cost

The practicality of an assessment method is also important. Whether an assessment tool is practical or not may depend on the resources and expertise available and its cost. Cost is an important characteristic of an assessment instrument if it is to be adopted for widespread use in a medical school. The cost of assessment instruments is difficult to determine and wide variations have been reported in the literature (Fowell *et al.*, 2000). Calculating the cost is even more complex than estimating the expenses related to the administration of a particular approach. Cost needs to include the start-up and continuing resources needed for development and implementation. The cost of an assessment should be considered in relation to the benefit to teaching and learning. Van der Vleuten (1996) has argued that an investment in a good assessment is also an investment in teaching and learning. What may be perceived initially as resource-intensive assessment approaches often turn out to be feasible in practice and, when the overall benefits are taken into consideration, they are deemed worth the expense.

These four factors—validity, reliability, impact on learning and practicality including cost—should be taken into account when instruments are chosen to assess the different learning outcomes.

Use of appropriate assessment tools

In the literature on assessment, numerous instruments are described which should find a place in the examiner's tool kit of assessment methods. This guide provides only a short overview of the different methods. It is beyond the scope of the guide to review or critique each of these assessment instruments in detail. The instruments can be considered in five categories—written assessments, clinical or practical assessments, observations, portfolios and other records of performance, and peer and self-assessment. Examples in each category are given in Table 3.

Table 3. Categories of assessment instruments with examples in each category.

Assessment category	Representative instruments
Written Assessments	Essay
	Short Answer Questions
	Completion Questions
	Multiple Choice Questions (MCQs)
	Extended Matching Items (EMIs)
	Modified Essay Questions (MEQs)
	Patient Management Problems (PMPs)
	Progress Test
	Dissertation
	Report
Clinical/Practical Assessments	Long Cases
	Practical Examination
	Spot Examination
	Objective Structured Clinical Examination (OSCE)
	Objective Structured Practical Examination (OSPE)
	Objective Structured Long Examination Record (OSLER)
	Group Objective Structured Clinical Examination (GOSCE)
Observation	Tutor's report
	Checklists
	Rating scales
	Patient report
Portfolio and Other Records of Performance	Logbooks
	Portfolios
	Procedural Logs
Peer and Self-Assessment	Peer report
	Self-report

A short description of each of the approaches is provided below.

Written assessments

Description. It is highly probable that readers are familiar with written examinations and assessments. They have been the most commonly used form of assessment in medical education. Essays and short-answer questions were formerly a prominent feature of examinations. There has been a general move in the past three decades, however, to more 'objective' or multiple-choice type of questions (MCQ). MCQs may be of the 'one from five' type where a statement is followed by 'distractors' that are included to discriminate the learner's ability to recall or apply factual information. In the UK, multiple true–false items where a common stem is followed by five statements, each of which may be true or false, were used. In the extended matching type a series of four or so items have to be matched with one from an extended list of options (e.g. appropriate investigations or treatment). Despite the increased emphasis on performance testing, the multiple-choice question format continues to play a part in low- and high-stake assessment. There is a trend to set MCQs in the context of patient scenarios. Modified essay questions and patient management problems (PMPs) offered potential advantages but are less popular now because of difficulties with question setting, with marking and with standardization.

The progress test was initially developed at Maastricht University School of Medicine about 20 years ago. It is a comprehensive test in MCQ format which samples a student's knowledge across the entire undergraduate medical education. It was designed so that, because of its depth, breadth and comprehensiveness, students could not prepare for the examination by cramming and memorizing of content. It is given four times per year to students in all years of the course and consists of approximately 250 items. In each administration a newly constructed test is prepared from the large item bank. As we would expect, students in the early years perform considerably less well than more advanced students nearing graduation. In other centres, it has been argued that short constructed-response answers are a more appropriate test of core knowledge and can be marked objectively. A progress test using this form of question has been described (Friedman Ben-David *et al.*, 2001a).

'Written' assessments are not necessarily limited to paper-and-pencil administration. They can be administered using computers. This provides a number of advantages and conveniences of administration, scoring, and sampling of students' mastery of content, including adapting the test to the needs of the individual student—so-called 'adaptive testing' (Hambleton, 1996). If medical students take their licensing exams on computers, as in the USA, they should be given opportunities for practice in their undergraduate programme.

Strengths and weaknesses. Written examinations, in particular the MCQ, have traditionally enjoyed high reliability. They can be conveniently administered but are time consuming to construct properly. They have been referred to as 'objective' as examiners do not make individual judgements about the quality of a student answer. MCQs, when constructed using a 'test blueprint', can sample a large domain of knowledge in an effective and efficient manner. Parallel forms of written examinations can be developed with equivalent content, quality and psychometric characteristics (Case, 2000). An over-reliance on the multiple-choice format to measure the recall of knowledge instead of higher level learning has resulted in a measure of disenchantment with multiple-choice testing. Critics of multiple-choice-type items have suggested that because they test discrete pieces of knowledge, and because they are designed to discriminate what a student knows or does not know, the questions tend to test 'trivial' facts or superficial knowledge of a subject rather than a deep understanding. A MCQ's cuing effects has been another criticism but the research does not appear to identify this as a major problem.

Impact on learning. The use of written assessment, particularly MCQs, has a significant impact on how students study and what they learn. To put it simply, students learn to pass the test. This can have unintended negative outcomes on their training to care for patients and to be competent and caring physicians. Written examinations are at the 'heart of the hidden curriculum' (Van der Vleuten, 1996). Students will study for the test rather than learn the information as an integrated whole. Who can blame their choice of increasing their chances to do well on the test! The effect of written exams on students is unpredictable and may have unanticipated consequences not predicted (Van der Vleuten, 1996). Faculty are also influenced by the use of written tests. Instruction is designed around what needs to be learned to pass the examinations and packaged according to when the examinations will be administered.

Practicality including cost. The design and construction of MCQ examinations is costly if it is to be done properly. However, MCQs can be shared between institutions and their administration and marking on a large scale is less expensive than other assessment approaches.

Clinical or practical assessments

Description. The multiple-station examination or the objective structured clinical exam (OSCE) is an assessment approach primarily used to measure clinical competence (Harden & Gleeson, 1979). Students are assessed at a number of stations on discrete focused activities that simulate aspects of clinical competence. Each student is exposed to the same stations and assessment. Examinees move from station to station in sequence on the same time schedule. The OSCE stations are designed to measure a number of predetermined clinical competences. Standardized patients (SPs), real patients or simulators may be used in the OSCE (Collins & Harden, 1998). OSCE stations may also incorporate the assessment of interpretation, non-patient skills and technical skills.

The OSCE is more than just a 'multi-station' exam. It is in essence a clinical or practical examination in which aspects of clinical competences are sampled to determine students' clinical skills and abilities related to their competence to practise medicine. The OSCE assesses performance and is concerned with what students 'can do' rather than what students 'know' as is the case with more traditional

assessment approaches. While the term OSCE refers to the class of examinations that use stations to measure performance, other names appear in the literature such as OSPE (objective structured practical examination). The OSPRE (objective structured performance-related exam) is used as a summative assessment in the police force in the UK. All of the examinations have similar characteristics and are essentially some modified special-purpose version of the original OSCE.

Standardized patients, simulations and models were introduced in medical education as the importance of an examination of the learner's clinical abilities was recognized. Written and observation approaches were inadequate as they did not accurately assess the required skills in a clinical encounter. Viewed simply, the clinical encounter has three variables—the learner, the patients, and the examiner. To assess a learner accurately, the patient and the examiner should be held constant across all the different learners to be assessed. In the literature there is a difference between what is meant by the terms 'simulated' and 'standardized' patients. A simulated patient is an individual who simulates the role of a real patient. A standardized patient is a patient, an actor or another individual who has undergone training to provide a realistic and consistent representation of a clinical encounter and who has been trained according to specific criteria to play the role of a patient with certain conditions. The terms standardized patients and simulated patients, however, are often used interchangeably and are often referred to as SPs. Simulated or standardized patients are particularly useful to help teach medical students history and physical examination skills. In the USA and Canada approximately 80% of medical schools use SPs in their medical educational programmes (Cushing, 2000) and we now have a great deal of information on how to use SPs effectively in OSCEs (Adamo, 2003). SPs have been used in assessment primarily in two ways: in the context of formal examinations such as an OSCE and in the day-to-day practice setting to assess the learner's or physician's performance in routine practice. Gorter *et al.* (2001) described, for example, how to introduce incognito standardized patients into outpatient clinics of specialists in rheumatology.

Simulations are approximations of reality and attempt to 'simulate' as near as possible a real clinical scenario or encounter. The use of simulations in medical education with paper-and-pencil encounters, computers or actors is widespread. One area where simulation is particularly prominent is in educating and assessing learners in history and physical diagnostic skills in the early as well as the later years of the undergraduate medical curriculum. Simulation is also used frequently in the surgical and anaesthesiology areas. Simulations have been used extensively to assess competence in clinical reasoning, patient examination, patient management and the performance of procedures. Simulations such as Harvey, the cardiac simulator, are now widely used in assessment (Issenberg *et al.*, 2001). Virtual reality offers the latest opportunity to simulate patient encounters and is an approach to watch out for in the future.

Strengths and weaknesses. The OSCE offers many advantages as a practical, reliable and valid tool to assess clinical competence. Norman (2002) has suggested that ''The objective structured clinical examination, with its multiple samples of performance has come to dominate performance assessment''. Generally, the greater the number of stations in the OSCE, the greater is its reliability and content validity. It has been stipulated that an OSCE with about 20 stations obtains the minimum reliability needed (ACGME and ABMS, 2000; Van der Vleuten & Swanson, 1990). Reliability is high when checklists are used with the standardized patient stations (Cushing, 2000). In a study of the use of the OSCE, the Accreditation Council for Graduate Medical Education found in paediatric certifying examinations a correlation between the OSCE and pre-certification examinations ranging between 0.59 and 0.71 (Carraccio & Englander, 2000). Given an adequate number of stations, the OSCE can be a reliable test with modest validity.

Probert *et al.* (2003) found that student performance in a traditional clinical final examination was not a good indicator of his/her rating as a junior doctor by a consultant. In contrast there was a positive relationship between student performance in a final year of OSCE with his/her rating as a junior doctor.

The advantage of using simulations for assessment purposes is that procedures which are normally difficult to assess can be assessed under conditions with high 'fidelity' without putting the learner or the patient at risk. In simulations, feedback to the learner can be given instantly so that the learner may be able to respond to that information and correct the behaviour. In Van der Vleuten and Swanson's review of the literature (1990) they found that in the use of multiple SPs very little measurement error was introduced when SPs were trained to play the same patient role. Vu *et al.* (1992) found that SPs were very good and consistent over time when recording checklist items. Reliability of SPs has been shown to be acceptable when there is adequate training and standardization (Tamblyn *et al.*, 1991).

In the assessment of students' clinical competence, each approach has its own advantages. One should not be limited to one assessment approach. Real patients, standardized or simulated patients, patient simulators or models may be the most appropriate approach depending on the particular skill being assessed and the context in which it is being assessed.

Impact on learning. Student participation in an OSCE has a positive impact on learning. One of the clear benefits is that the students' attention is focused on the acquisition of clinical skills. Students are examined in settings that better approximate the skills and competences that will be required when they practise medicine. The nature of the OSCE is to sample and simulate the examination and management of a patient. The OSCE provides formative evaluation as the student is participating in it. Participation in an OSCE was found to improve competence at subsequent stations and improve the quality of the learning experiences for examinees (Carraccio & Englander, 2000).

The OSCE also has a potentially negative impact on learning. An OSCE contains multiple stations that sample aspects of clinical medicine. The student may prepare for the examination by compartmentalizing the skills and not completely understanding the connection and flow of the skills.

The use of SPs, simulations and manikins for assessment allow practical and clinical competences to be assessed in a formal examination and focus the student's attention on these competences.

Practicality including cost. The cost of implementing an OSCE varies greatly from centre to centre. The time and resources required to administer an OSCE are the two major costs. Estimated costs have been hard to pin down, and have been reported as ranging from US$7 per student per station to between US$54 and US$496 per student to participate in an OSCE (Carraccio & Englander, 2000). Contributing to the costs are payment for standardized patients, required facilities and administrative and examiners' time. While these are charges in some centres, particularly in the USA, in others only marginal costs are incurred, standardized patients provide their services free and no charge is being made for the use of the facilities where the examination is held. Where an OSCE is mounted within an institution or medical school, costs relating to staff and facilities may be provided at no additional cost. Where the OSCE is administered in a special facility designed for the purpose, the costs may be significant. The costs of the preparation and planning of the examination may be considerable. Maintaining exam security is another potential cost.

The recruitment, training and scheduling of standardized patients may be for some centres one of the most expensive and most variable aspects contributing to the overall cost of standardized patients. The start-up and continuing costs of a standardized patient programme can seem overwhelming and often discourage institutions from investing in such an activity. Approaches vary from programme to programme. There is also a considerable investment in staff time and resources, which may include special facilities for the SPs to be examined by learners. Recruitment of SPs is another time-consuming task. It has been estimated that the time required to train SPs ranges from one to 15 hours depending on the clinical scenario being simulated. Costs also vary according to whether laypersons, actors or volunteers are used.

Observation

Description. A time-tested and well-used method of assessing students in medicine is the observation of the student by clinical faculty members during a clinical attachment. In the United States it is the predominant method used to assess students' performance during the third- and fourth-year clinical clerkships. Recently the use of observation has been gaining popularity during the first two years.

Faculty usually record their assessment of students on a checklist or rating scale. The student can be marked or rated against each of the expected learning outcomes, e.g. communication skills and attitudes. A global rating of the student's overall competence may also be recorded. For practical procedures, the faculty may simply 'check' whether the procedure was performed or not. It is especially helpful if the examiner includes narrative comments with the ratings.

Strengths and weaknesses. One of the major issues in the assessment of students by faculty is the lack of reliability of faculty ratings. This may be due to the fact that staff have had insufficient contact with the student to make an informed judgement of the student's abilities. The consistency of ratings may be improved with training of the staff. Scores may be biased because of different standards applied by untrained raters who do not want to include the ends of the rating scales in their evaluations and who do not have the ability to discriminate between the different categories on the rating scale. Ratings for the acquisition of knowledge appear to be more reliable than ratings for interpersonal or communication skills (ACGME and ABMS, 2000).

We have described the rating of students during clinical attachments. Students' performance and abilities may also be rated by staff in other situations, such as tutorials or PBL sessions.

Impact on learning. If a student's performance on learning outcomes, not assessed with other instruments, is rated over a period of time, the student's attention is directed to these outcomes that might otherwise have been ignored. Such assessments can also be a powerful source of feedback to students. If the ratings are used for summative purposes there is a danger that the students will be encouraged to hide their weaknesses and deficiencies rather than reveal them to the teacher with a view to rectifying them.

Practicality including cost. The time and commitment needed for the evaluation of students over time by faculty should not be underestimated. The bottom line is that good assessment practice requires staff time, training and motivation.

Portfolios and other reports of performance

Description. A portfolio is a collection of material made by a professional or a learner that documents his or her achievements and includes a reflection on those achievements. Portfolio-based learning can be defined as "the collection of evidence that learning has taken place, usually set within some agreed objectives or a negotiated set of learning activities" (Snadden, 1999). They contain material collected by the learner over time and include a critical review and analysis by the student of the material. Portfolios can contain just about anything. The following are typical elements of a portfolio: critical incidents or events, a reflective journal or diary, tutorials and learning plans, clinical experiences, exam preparation materials, recorded consultations, audits and project work, critical reviews of articles, and management material (Snadden, 1999). Portfolios used for assessment purposes (both formative and summative) need to be written primarily so they can be reviewed. Issues of public and private viewing need to be made explicit and clear from the beginning. The learner should control who has access and can see her/his portfolio because of the potential personal nature of reflection that may occur. Assessment of portfolios concentrates on whether the learner has demonstrated in the portfolio that the learning outcomes have been achieved. Portfolios are very useful to document that learners have achieved the desired learning outcomes (Davis *et al.*, 2001, Friedman Ben-David *et al.*, 2001b).

Logbooks, like portfolios, document the student's experiences. However, they are usually more limited in scope than portfolios and are focused on data collected in a specific area or activity. Reflection is not normally part of logs. At least three kinds of logs have been documented—procedural, operative and case logs (ACGME and ABMS, 2000). Procedural logs usually document how many and when procedures were performed by the learner. Operative logs are similar but document what was done and when. Case logs

record which patients were seen and with what diseases, within a given time period.

Strengths and weaknesses. Portfolios are a valuable instrument for inclusion in the examiner's toolkit, if for no other reason than that they assess learning outcomes such as critical thinking and self-assessment, which are not easily assessed using other instruments. They also provide a record of the student's performance over a period of time and are not just a snapshot at one specific point in time. The reliability of portfolios is at least in part due to the ability of raters to agree on standards and criteria for the content and assessment of portfolios. Their reliability is enhanced by the triangulation of evidence from a number of sources (Friedman Ben-David *et al.*, 2001a). The validity of portfolios will be determined by the extent to which they document accurately those experiences that are indicators of the mastery of the desired learning outcomes.

Logbooks are limited in their assessment powers. Logbooks involve simply recording whether something has been performed or not. They do not document competence or quality. The literature is weak in determining whether logbooks are accurately completed by the learners (ACGME and ABMS, 2000). The reporting of validity and reliability data for logbooks is unavailable.

Impact on learning. Portfolio assessments have a positive impact on the learning environment because they document what the learner has done and ask the learner to reflect on what she/he has accomplished. They ask the learner to put together large amounts of disparate information into a whole that tells a story about the learner's accomplishments. The portfolio is more than an accumulation of what the learner has done. It is a story of the development of the learner and, seen as such, can greatly contribute to the positive nature of how a student learns and studies.

Logbooks are often considered by learners to be boring or repetitive. Yet, viewed from the preceptor's point of view, they can paint a picture of what the learner has or has not done. The learner also has an opportunity to see what he or she has accomplished by reviewing the documented list of tasks contained in the logbook.

Practicality including cost. It goes without saying that these assessment approaches require staff time and resources. The portfolio assessment has its own associated costs, especially those related to the external review and marking of the portfolios. There is cost associated with developing electronic portfolios or logbooks. These resource investments can be justified, however, in the light of the valuable insight gained into the learner's abilities and competence to practise medicine.

Peer and self-assessment

Description. Peer and self-assessment focuses on an important consideration in assessment: who should do the evaluating? Most evaluation methods are designed by faculty who 'naturally' take on the responsibility for assessing learners. However, important information can be gained if we also ask the learner's peers and the learners themselves what they are able to do and how comfortably they are performing a task.

Peer and self-assessment may carry less weight than ratings by trained faculty examiners. It is likely, therefore, that peer and self-assessment will be used in conjunction with ratings by faculty and other trained health professionals. Checklists and rating scales can be used. Peers have the advantage of observing each other performing the tasks and procedures that are being learned under real clinical conditions (e.g. residents are more likely than the attending physician to observe their peer's performance). Peers can be more discriminating than faculty in their evaluations because of their increased exposure to the observed performance. Peer ratings have particularly been used in the area of the assessment of attitudes and communication skills among medical students. One of the concerns in the education of medical students is: 'Will the student have the ability to know when he/she doesn't know, and will he/she seek help?' In other words, the learner needs to be able to self-evaluate and take the appropriate action from that assessment. The use of self-assessment shows promise in understanding how to assess this important attribute of lifelong learning and self-discriminating abilities.

Strengths and weaknesses. Studies have shown that peer evaluations correlate highly with faculty evaluations of the rating of the same behaviours and skills (Gray, 1996). Self-assessment correlates moderately with the rating of a trained examiner. It is reported that mean ratings by self-raters tend to be lower (more critical) than those of trained examiners (Cushing, 2000). One of the difficulties in the use of peer and self-assessment approaches is the training of the raters. It has been found that the provision of benchmarks helps to improve the reliability of such ratings. A benchmark is a standard or expected level of performance with an example that helps the rater standardize his/her rating compared with those of others. Reports of the use of peer and self-assessment suggest that this approach to assessment is little used in the majority of medical education experiences. Where it has been used is mainly to assess communication skills in small groups and other similar learning settings, for example PBL (Fowell *et al.*, 2000). The use of peer and self-assessment are areas that need further study and development.

Impact on learning. The use of peer and self-assessment can have a profound impact on the students' educational programme. On the positive side, their use has the power to transform the perceived nature of an examination, especially in the clinical years, and to develop in the student skills of self-appraisal. On the negative side, if done poorly, their use could cause mistrust, suspicion and peer rivalries that would be devastating to a medical education programme. It is difficult to predict the results of a widespread use of such approaches. Which will it be: negative or positive? It is up to the institution and the climate of trust and acceptance of change to determine the final outcome.

Practicality including cost. The main cost and time commitment is the training of peers and of the learners themselves to be reliable and accurate raters.

Assessment of the full range of learning outcomes

An overview

Miller (1990) proposed a pyramid of learning with increasing professional authenticity, starting with the learner's cognition and moving towards a focus on the learner's behaviour. Van der Vleuten (2002) linked a hierarchy of assessment approaches of increasing authenticity with the Miller Pyramid. This model is useful with regard to assessment in so far as different assessment instruments are appropriate at each level of the pyramid.

Figure 3 shows the Miller pyramid with each of the four levels matched with one or more of the assessment approaches described above. Written assessment is the predominant instrument at the 'know' and 'know how' levels, clinical assessment and the OSCE at the 'shows how' level and observation, portfolios and logbooks at the 'does' level.

Assessment and the learning outcomes

This section of the guide looks at assessment instruments with regard to their appropriateness as tools to measure the attainment by the students of the different learning outcomes. It describes 12 learning outcomes and provides a commentary on their assessment.

Table 4 lists for each of the 12 learning outcomes the most appropriate assessment approaches. All of the methods can be used to a greater or lesser extent to assess each of the outcomes. The methods are arranged according to the authors' estimation of their relative importance for each outcome, with the most important or useful instrument for each learning outcome listed first.

Knowledge is embedded in each of the learning outcomes, for example knowledge about clinical theory, knowledge about management, knowledge about ethical principles and knowledge about the healthcare team and teamwork. Thus instruments used to assess knowledge may be appropriate in relation to all 12 outcomes. With regard to the technical skills of the student, as exemplified in outcomes one to seven, the assessment should include the demonstration of the relevant skill. Clinical and practical assessment methods are therefore particularly important. Attitudes, decision making and professionalism can be demonstrated most effectively through instruments designed to assess the student's or doctor's performance in practice in the clinical context. The four levels of the Miller pyramid are relevant to all 12 learning outcomes. Figure 4 lists for each of the learning outcomes what is reported as the most important level of the Miller pyramid for the purposes of assessment.

The assessment of technical competences ('What the doctor is able to do')

Learning Outcome 1: Competence in clinical skills. The doctor must be able to take a complete and focused history, perform an appropriate physical exam, interpret findings, and formulate an action plan to characterize the problem and reach a diagnosis.

To allow a judgement to be made with regard to a student's competence in clinical skills such as taking a history or examining a patient's abdomen, the student needs to demonstrate that he or she can perform the necessary clinical skill. The student must 'show how' he/she takes a history, examines a patient as appropriate, interprets the findings and formulates an action plan in order to reach a final diagnosis. Clinical examinations, particularly the OSCE, are appropriate for this purpose and may provide sufficient information on which the student's abilities with regard to these outcomes can be assessed.

It may be possible, however, to obtain information about the achievement of these outcomes from reports of observers such as clinical attachment supervisors or from student entries in a diary or logbook presented as part of a portfolio. Some teachers choose to ignore such additional sources of information about the student's clinical skills because of the potential unreliability of the data or the cost of collecting and analysing it.

Learning Outcome 2: Competence in practical procedures. The doctor should be able to undertake a range of procedures on a patient for diagnostic or therapeutic purposes. This usually involves using an instrument or some device, e.g. suturing a wound or catheterization.

Competence in some procedures can be assessed in the same way as clinical skills, with a practical examination such as an OSCE. This is not possible or desirable for all procedures. A record of the student's performance of the required procedure in practice, certified by a member of staff, may be an important part of a student portfolio or logbook which is assessed by the examiners. Information about the student's competence in practical procedures may also be

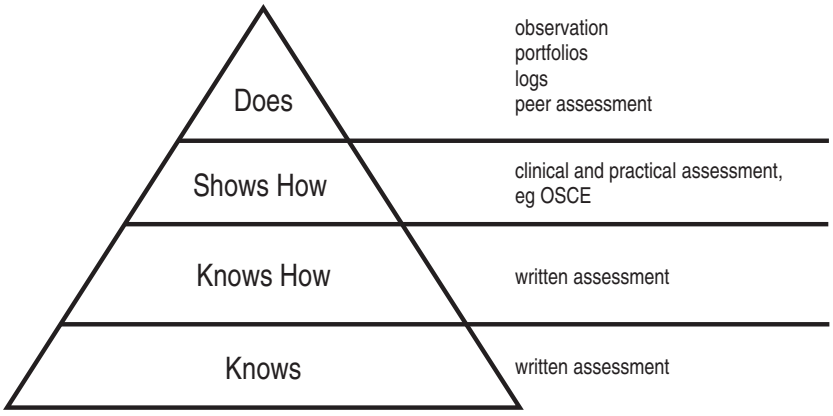

Figure 3. The learning assessment pyramid.

Table 4. Recommended assessment methods for the 12 learning outcomes of a competent and reflective physician.

Learning outcome	Assessment methods
What the doctor is able to do	
1 Clinical Skills	OSCE
	Observation
	Logbooks
	Written examination
2 Practical Procedures	OSCE
	Portfolios and logbooks
	Observation
	Written examination
3 Patient Investigation	Written examination
	OSCE
	Observation
	Portfolio
4 Patient Management	Written examination
	OSCE
	Observation
	Portfolios
5 Health Promotion and Disease Prevention	OSCE
	Portfolios
	Observation
	Written Assessment
6 Communication	OSCE
	Observation
	Peer/self assessment
	Portfolio
7 Information Management Skills	Portfolio
	OSCE
	Observation
	Written examination
How doctors approach their practice	
8 Principles of Social, Basic and Clinical Sciences	Written examination
	Portfolios
	Observation
	OSCE
9 Attitudes, Ethics and Legal Responsibilities	Observation
	Portfolio
	Peer/self assessment
	OSCE
	Written examination
10 Decision Making, Clinical Reasoning and Judgement	Portfolio
	Observation
	Written examination
	OSCE
	Peer/self assessment
Doctors as professionals	
11 Role as a Professional	Observation
	Peer/self assessment
	Portfolio
	OSCE
	Written examination
12 Personal Development	Portfolio
	Observation
	Peer/self assessment
	OSCE
	Written examination

obtained from staff responsible for supervising a student during an attachment.

Learning Outcome 3: Competence to investigate a patient. The doctor should be competent to arrange appropriate investigations for a patient and, where appropriate, interpret these. The investigations are carried out on the patient or on samples of fluid or tissue taken from the patient. The investigations are usually carried out by personnel trained for the purpose, e.g. a clinical biochemist or radiographer, but may in some instances be carried out by the doctor.

Written assessment instruments can be used to provide the student with the opportunity to demonstrate that he/she knows how to investigate a patient and that he/she can, where appropriate, interpret the results of patient investigation. Students may be observed doing this in an OSCE. Their competence over time can be assessed by clinical supervisors during clinical attachments and in a portfolio.

Learning Outcome 4: Competence in patient management. The doctor is competent to identify appropriate treatment for the patient and to deliver this personally or to refer the patient to the appropriate colleague for treatment. It includes interventions such as surgery and drug therapy and in contexts for care such as acute care and rehabilitation.

The examiner, using a combined approach to assessment with written instruments, OSCEs, observation and portfolios, is able to assess whether the student can successfully integrate knowledge and apply it competently to the management of a patient. Both indirect (written) and direct (OSCE) assessment approaches work together to understand the abilities of the learner.

Learning Outcome 5: Competence in health promotion and disease prevention. The doctor recognizes threats to the health of individuals or communities at risk. The doctor is able to implement, where appropriate, the basic principles of disease prevention and health promotion.

The assessment of a learner's ability to practise medicine with a population medicine perspective aimed at disease prevention and health promotion requires the application of assessment instruments in all three domains: cognitive, skill and affective. The assessment should occur in either a simulated or preferably a practice setting because of the need to assess a more complex process encompassing the broader perspective of health and how to promote it. OSCEs and portfolios are particularly useful. The knowledge underpinning health promotion and disease prevention and its application can be assessed with written examinations.

Learning Outcome 6: Competence in communication. The doctor is proficient in a range of communication skills including written and oral, both face-to-face and by telephone. He or she communicates effectively with patients, relatives of patients, the public and colleagues.

The assessment of communication skills has not in the past attracted as much attention as it deserves. It is, however, one of the competences that are critical to the care of the patient. Good care is only as good as the ability to communicate clearly and accurately with fellow health workers and other professionals, and with patients and their families. Communication skills are best assessed under 'real'

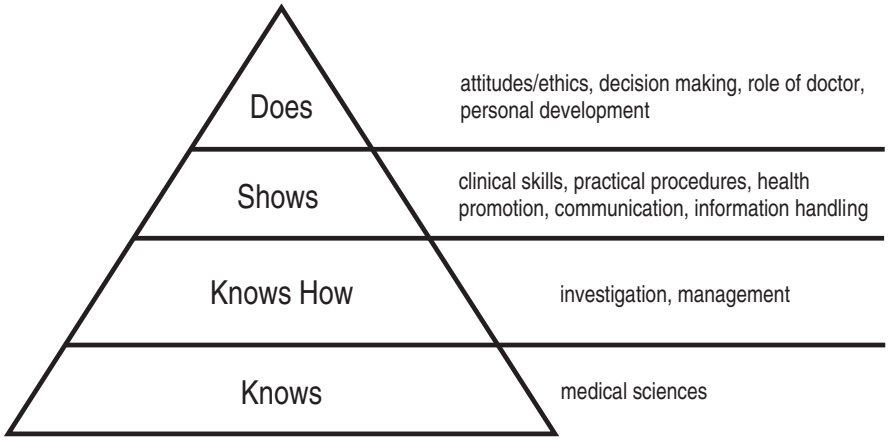

Figure 4. The 12 learning outcomes matched for assessment purposes against the most appropriate level of the Miller Pyramid.

or simulated conditions in the care of patients. The assessment of communication skills can occur continuously in the curriculum from the early years of medical school and in the student's progress in the area recorded. The assessment of communication skills is a sensitive area because of the personal nature of learners' communication abilities, styles, background, etc. This assessment has a profound impact on the learning environment that should not be underestimated. Careful planning has to occur in the assessment of communication skills so as not to interrupt or intrusively interfere with the patient–physician relationship. The assessment of communication skills should be high on the list of assessed competences of any medical school. A good measure of this competence can tell one much about learners' potential success and their ability to become competent and caring physicians. OSCE's, observation, peer assessment and portfolios have been used effectively to assess different aspects of communication skills.

Learning Outcome 7: Competence in handling and retrieval of information. The doctor is competent in retrieving, recording and analysing information using a range of methods including computers.

Students should be assessed as to their competence to retrieve, record and analyse information obtained from a range of sources including information technology resources. This learning outcome can be assessed by the student, documenting her/his experience in a portfolio or demonstrating these skills in an OSCE. The students' competence can also be assessed over time by observation.

Assessment of intellectual, emotional, analytical and creative competences ('How the doctor approaches his/her practice')

Learning Outcome 8: Approach practice with an understanding of basic and clinical sciences. Doctors should understand the basic, clinical and social sciences that underpin the practice of medicine. This ensures that they not only have the technical competence outlined in outcomes 1 to 7 but that they also have an understanding of what they are doing, and why they are doing it. This includes an awareness of the psychosocial dimensions of medicine.

The assessment of an understanding of basic and clinical sciences is heavily concentrated in the cognitive domain. The classic approach to the assessment of the 'sciences' in medicine is through written tests. Portfolios and logbooks are also helpful in that they require students to reflect on the relationship of what they know to the application of what they do in a care setting.

Learning Outcome 9: Approach practice with appropriate attitudes, ethical stance and legal responsibilities. Doctors adopt appropriate attitudes, ethical behaviour and legal approaches to the practice of medicine. This includes issues relating to informed consent, confidentiality and the practice of medicine in a multicultural society.

The assessment of attitudes, ethical understanding and legal responsibilities is a relatively new area of assessment in medical education. It is now recognized, however, to be a key aspect of competence for the future physician. ''Medical schools need to be satisfied that each of their students has reached a minimum standard of ethical competence, just as they need to be satisfied with the clinical competence of their graduates'' (Wong & Cheung, 2003). Attitudes, ethical understanding and legal responsibilities can be assessed in a variety of ways. The assessment of attitudes, ethics and legal responsibilities is high on the learning pyramid and occurs best in the 'real' setting during care giving. It follows that these outcomes are best assessed directly over time by staff observing the student's behaviour. Peer assessment and portfolios can be powerful assessment tools in this area. Written assessment approaches, computer simulations and OSCEs can be used as a secondary measure to assess the application of students' attitudes, ethics and understanding of their legal responsibilities in the simulated or practice setting. The use of multiple instruments and observations based on multiple samples of behaviour is particularly important in this difficult area.

Caution needs to be exercised when students' behaviours are assessed in practice to ensure that the assessment itself does not interfere with the student's behaviour. Students may unknowingly alter their caring behaviour with patients because of a fear of being criticized. Caution also needs to be exercised in the use of feedback to learners in this area. Learners may be particularly sensitive to criticism.

Learning Outcome 10: Approach practice with appropriate decision making, clinical reasoning and judgement. Doctors apply clinical judgement and evidence-based medicine to their practice. They understand research and statistical methods. They can cope with uncertainty and ambiguity. Medicine requires, in some cases, instant recognition, response and unreflective action, and at other times deliberate analysis and decisions and actions following a period of reflection and deliberation. This outcome also recognizes the creative element in problem solving that can be important in medical practice.

Problem solving is a higher-order cognitive activity. Its assessment needs to occur in the setting where problem solving is applied and justified. Decision-making skills and clinical reasoning are difficult to assess because there are no generic or effective universal approaches to their assessment. This is due to several factors, the most notable of which is that clinical reasoning approaches are case dependent. In other words, different decision-making approaches are applied related to the nature of the problem. It appears that the difference between novice and expert physicians in their clinical reasoning skills is attributed to 'problem scripts' that are developed over time based on accumulated experience. The assessment of clinical reasoning on the part of the medical student can best be thought of as a reflective process with demonstrated accumulated evidence. Patient management problems (PMPs) used in the early 1970s were difficult to score and it was difficult to determine the exact 'decision path' that a learner was following. Extended matching items (EMIs) get at the knowledge aspect of clinical decision making and how students use the knowledge in this context. Portfolios can be used conveniently to assess the student's reflective practice. Students demonstrate that they can reflect on what they have accomplished, and examine and explain their problem-solving approaches on a case-by-case basis. Problem solving and decision making can also be assessed in the OSCE. An assessment approach, the Diagnostic Thinking Inventory (Bordage *et al.*, 1999), was designed specifically to assess a student's problem-solving abilities.

Assessment of personal competences ('The doctor as a professional')

Learning Outcome 11: Physicians should have an appreciation (understanding) of the doctor's role in the health service. Doctors should understand the healthcare system within which they are practising and the roles of other professionals within the system. They should appreciate the role of the doctor as physician, teacher, manager and researcher. There should be a willingness by the doctor to contribute to research even in a modest way and to build up the evidence base for medical practice. This outcome also recognizes that most doctors have some management and teaching responsibility.

Many of these competences are best assessed by observing directly or indirectly the student's or the doctor's behaviour using observations, ratings, reports or portfolios. The OSCE can assess the learning outcomes in a simulated situation. For example, a healthcare team including doctors and nurses can be assessed as a team in a group OSCE. Knowledge components can be assessed in written tests.

Learning Outcome 12: Physicians should have an aptitude for personal development. The doctor should be a self-learner and should be able to assess his or her own performance. The doctor has to take responsibility for his or her own personal and professional development, including personal health and career development. These are important learning outcomes.

Their assessment is, however, not easy. One of the most important qualities a physician can possess is the ability to judge his or her personal limits or abilities. A specific assessment instrument to determine whether a student is a self-directed and lifelong learner has not yet been devised. These skills and behaviours are best directly observed and assessed in the clinical phase of a medical student's education by a clinical supervisor. The assessment of professionalism may ultimately best be done by oneself and one's peers. Another approach which provides a glimpse into professionalism and self-directedness is the use of portfolios for assessment purposes. The portfolio provides the learner with the opportunity to demonstrate evidence of self-reflection on the accumulated evidence of her/his experiences and accomplishments during her/his medical education.

Discussion and conclusions

The development in medical education institutions of appropriate approaches to assessment underpinned by a related philosophy has lagged behind developments that have occurred in other aspects of the curriculum. The implications of this may be serious. Implementation of a new curriculum without changes to the approach to assessment may result in little or no change at all. More attention must be paid to assessment and the underlying principles (Brown *et al.*, 1996; Fowell *et al.*, 2000). Faculty and staff at each institution should spend time developing a cohesive assessment philosophy and engaging in a staff development programme. Engagement with the assessment process and ensuring that students achieve the required learning outcomes is an important role of a medical teacher (Harden & Crosby, 2000).

This guide outlines what we are looking for in an assessment instrument and describes the wide range of tools that are available to assess the required range of learning outcomes. It emphasizes the move from assessing knowledge and technical competence to more complex learning outcomes such as attitudes, teamwork and professionalism. Written assessments, OSCEs, standardized patients, simulations and models, observation, checklists and rating scales, portfolios and logs, and peer and self-assessment all deserve attention. How they fit and complement one another in the assessment of the required competences is as important as understanding the details of each.

It is unlikely that one assessment instrument can address all of the learning outcomes. In general, what is required is a focus on the construction of test blueprints to adequately sample the learning outcomes to be assessed. It is necessary to choose assessment approaches that will do the job. New assessment instruments have been developed and introduced, and can be used to assess the range of learning outcomes. More traditional instruments such as MCQs that have been used for decades have dominated the assessment process. We are likely to see in the future a move away from selected-

response questions as in MCQs to constructed-response questions. There will also be a move to the assessment of learning outcomes which are not currently tested. Institutions need a basic package of assessment methods that will get the job done. The basic assessment package should consist of some sort of written assessment, for example, constructed-response questions (CRQs), and/or extended-matching items. The toolkit should also include a performance assessment such as the OSCE. Finally some measure of the students over time to assess other outcomes such as critical thinking and self-assessment is necessary. Portfolios can be used for this purpose. This basic package is not meant to be exhaustive, indeed institutions are encouraged to develop additional assessment strategies that meet the unique and particular needs of their own settings.

What is needed is a move to assess what learners do in practice and how they apply their knowledge of basic and clinical sciences to the care of patients. While many schools have moved to the assessment of competence in stimulated situations, only a few have placed a major emphasis on the assessment of performance in clinical practice through direct observational approaches or through the use of indirect measures such as portfolios.

Qualitative assessment approaches have also been under-used in medical education. Qualitative assessment has been associated wrongly with subjective assessment. Evidence-based literature exists that outlines the principles of good qualitative assessment (MacQueen *et al.*, 1998; Murphy *et al.*, 1998). Qualitative assessment methods provide the assessor with a sense of the environment and conditions in which the learning and its practice take place. It allows the assessor to better understand not just what students know or do not know, but what they do with that knowledge in the real settings that require the application of the learner's knowledge, attitudes and skills. Three principles apply in the use of qualitative assessment: triangulation, frequent and continuous assessment, and training of the evaluators. Triangulation means measuring the same phenomenon from different angles or vantage points using a variety of measurement strategies. In other words, assessors 'triangulate' on the phenomena being measured. Cushing (2000) described triangulation as a means whereby data are acquired from different sources and found to be consistent so that they can corroborate validity and reliability. Multiple measures need to be employed because of the complexity of the assessment of higher-order application of knowledge, attitudes and skills in the care of patients. The second principle is that assessment should be continuous and frequent. In order to sample the learner's abilities and attributes, one cannot rely on a single measure at a single point in time. Such samples are inherently unreliable, and their validity can be questioned. As acquired knowledge is applied in the clinical setting to the actual practice of medicine, the application of that knowledge becomes more case or situation specific. In order to adequately sample the students' competence, more frequent assessment needs to occur. Assessments that occur beforehand are helpful to learners so that they can assess where they are and what they need to do to improve. Learners are developing and changing with each new learning experience. It seems only logical that a true assessment of their competence would be the latest assessment. The third principle is that assessors need to be trained. Nowhere else is this more true than in the clinical setting. In order for faculty to assess learners' abilities they need to be trained how to do it. A specific example is the need for inter-rater agreement on standards and on marking scale scores. The practice of evidence-based assessment means that the instruments used and the application of those instruments will be valid and reliable. The need for staff and faculty development as assessors underpins the ability to achieve the desired standards.

We have referred to the assessment of each learning outcome and to the choice of the most appropriate tool to be used for the purpose. While an assessment of each competence or learning outcome is necessary, an indication of a student's strength and weakness across all competences is the ultimate goal. This can be presented as a profile for a student which shows his/her levels of attainment for each of the competences. Figure 5 illustrates a hypothetical profile for a student. In the example provided, the student reaches the required standard in all of the outcomes with the exception of those relating to attitudes and professionalism. Such a profile

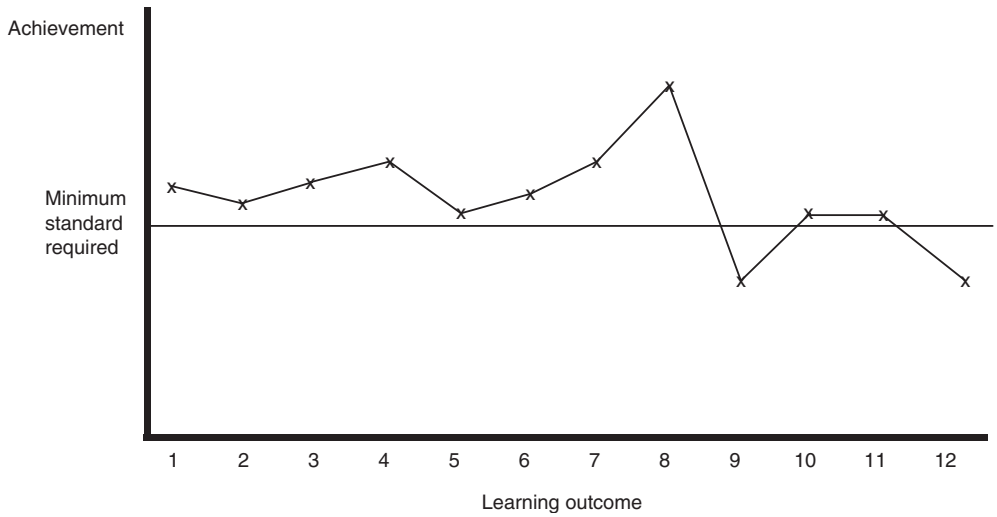

Figure 5. Profile for a student.

can prove to be very useful to the educational director, to the student and to curriculum developers.

In response to increasing public demands for a greater measure of accountability for the physicians we educate, rapid progress needs to be made in designing competence-based curricula and assessing students in increasingly realistic ways to show they can practise medicine. In this guide we demonstrate how to do a better job of assessing our students: we need an understanding of the assessment process and knowledge of the tools available and how they can be used to assess the range of learning outcomes. This model of assessment of learning outcomes needs to be applied across the different phases of medical education from undergraduate medical education to postgraduate education and continuing professional development. We need more studies of how professional competence is measured and how those measurement data are applied so that desired changes in behaviour are achieved. We have an opportunity to work across the medical education continuum to significantly improve what we do educationally, to have an impact on the future of the practice of medicine, and to guarantee to the public the competence of our physician workforce. In the end we will look back and be able to confidently educate the 'competent and reflective' physician.

Notes on contributors

James M. Shumway is Associate Dean for Medical Education and Professor of Internal Medicine and Pediatrics at West Virginia University School of Medicine

Ronald M. Harden is Education Director of the International Virtual Medical School (IVIMEDS) and former Director of the Centre for Medical Education, University of Dundee.

References

AAMC (1998) *Report I: Learning Objectives for Medical Student Education*, Medical Schools Objectives Project, January (Washington, AAMC).

ACGME (2003) Accreditation Council for Graduate Medical Education [website: http://www.acgme.org].

ACGME and ABMS (2000) *Toolbox of Assessment Methods*, A Product of the Joint Initiative, Accreditation Council for Graduate Medical Education (ACGME) Outcomes Project, American Board of Medical Specialties (ABMS), Version 1.1, September 2000 [Copyright 2000 Accreditation Council for Graduate Medical Education and American Board of Medical Specialties].

Adamo, G. (2003) Simulated and standardized patients in OSCEs achievements and challenges 1992–2003 *Medical Teacher*, 5(3), pp. 262–270.

Atkinson, R.L., Atkinson, R.C., Smith, E.E. & Bem, D.J. (1993) *Introduction to Psychology*, 11th edn (New York, Harcourt Brace Jovanovich College Publishers).

Bordage, G., Grant, J. & Marsden, P. (1999) Quantitative assessment of diagnostic ability, *Medical Education*, 24, pp. 413–425.

Brown, S., Race, P. & Smith, B. (1996) *500 Tips on Assessment* (London, Kogan Page.)

CanMEDS (2000) Extract from the CanMEDS 2000 Project Societal Needs Working Group Report, *Medical Teacher*, 22(6), pp. 549–554.

Carraccio, C. & Englander, R. (2000) The objective structured clinical examination: a step in the direction of competency-based evaluation, *Archives of Pediatric Adolescent Medicine*, 154, pp. 736–741.

Case, S. (2000) Developments in written assessment, oral presentation given at the 9th International Ottawa Conference on Medical Education, Cape Town, South Africa, 28 February–3 March [http://www.educ.unimaas.nl/ottawa/].

Collins, J.P. & Harden, R.M. (1998) AMEE Education Guide No. 13: The use of real patients, simulated patients and simulators in clinical examinations, *Medical Teacher* 20(6), pp. 508–521.

Cushing, A. (2000) Developments in attitude and professional behaviour assessment, oral presentation given at the 9th International Ottawa Conference on Medical Education, Cape Town, South Africa, 28 February–3 March [http://www.educ.unimaas.nl/ottawa/].

Cushing, A. (2002) Assessment of Non-Cognitive Factors, in: G.R. Norman (Ed.) *International Handbook of Research in Medical Education*, Ch. 22 (Dordrecht: Kluwer).

Davis, M.H., Friedman, Ben-David, M., Harden, R.M., Howie, P., Ker, J., McGhee, C., Pippard, M.J. & Snadden, D. (2001) Portfolio assessment in medical students' final examinations, *Medical Teacher*, 23, pp. 357–366.

Fowell, S.L., Maudsley, G., Maguire, P., Leinster, S.J. & Bligh, J. (2000) Report of findings: student assessment in undergraduate medical education in the United Kingdom 1998, *Medical Education*, 34(Suppl. 1, September), pp. 1–78.

Friedman Ben-David, M. (1999) AMEE Guide No 14: Outcome-based education, part 3: Assessment in outcome-based education, *Medical Teacher*, 21(1), pp. 23–25.

Friedman Ben-David, M., Hunter, I. & Harden, R.M. (2001a) Introduction of a progress test as a tool for defining core undergraduate curriculum, *Educacion Medica XV Congreso de la Sociedad Nacional de Educacion Medica*, 4(3), pp. 97–99.

Friedman Ben-David, M., Davis, M.H., Harden, R.M., Howie, W., Ker, J., McGhee, C., Pippard, M.J. & Snadden, D. (2001b) AMEE Education Guide No. 24: Portfolios as a method of student assessment, *Medical Teacher*, 23(6), pp. 535–552.

General Medical Council (1993) *Tomorrow's Doctors: Recommendations on Undergraduate Medical Education* (London, GMC).

General Medical Council (2002) *Recommendations on Undergraduate Medical Education* (London, GMC).

Goetz, E.T., Alexander, P.A. & Ash, M.J. (1992) *Educational Psychology A Classroom Perspective* (New York, Maxwell Macmillan International).

Gorter, S.L., Rethans, J.J., Scherpbier, A.J.J.A., Van Der Linden, S. *et al*. (2001) How to introduce incognito standardized patients into outpatient clinics of specialists in rheumatology, *Medical Teacher*, 23(2), pp. 138–144.

Gray, J. (1996) Global rating scales in residency education, *Academic Medicine* 71, pp. S55–63.

Hambleton, R.K. (1996). Advances in assessment models, methods, and practices, in: D.C. Berliner & R.C. Calfee (Eds) *Handbook of Educational Psychology*, pp. 899–925 (New York, Simon & Schuster Macmillan).

Harden, R.M. (2002) Developments in outcome-based education, *Medical Teacher*, 24(2), pp. 117–120.

Harden, R.M. & Crosby, J.R. (2000) AMEE Education Guide No. 20: The good teacher is more than a lecturer—the twelve roles of the teacher, *Medical Teacher*, 22(4), pp. 334–347.

Harden, R.M., Crosby, J.R. & Davis, M.H. (1999a) An introduction to outcome-based education, *Medical Teacher*, 21(2), pp. 7–14.

Harden, R.M., Crosby, J.R., Davis, M.H. & Friedman, M. (1999b). From competency to meta-competency: a model for the specification of learning outcomes, *Medical Teacher*, 21(6), pp. 546–552.

Harden, R.M. & Gleeson, F.A. (1979) ASME Medical Education Booklet No. 8: Assessment of clinical competence using an objective structured clinical examination (OSCE), *Medical Education*, 13, pp. 41–54.

IIME Core Committee (2002) Global minimum essential requirements in medical education, *Medical Teacher*, 24(2), pp. 130–135.

Issenberg, B., Gordon, M.S., Gordon, D.L., Safford, R.E. & Hart, I.R. (2001) Simulation and new learning technologies, *Medical Teacher*, 23(1), pp. 16–23.

Kassebaum, D.G., Eaglen, R.H. & Cutler, E.R. (1997). The objectives of medical education: reflections in the accreditation looking glass, *Academic Medicine*, 72(7), pp. 648–656.

LCME (2003) Liaison Committee on Medical Education [website: http://www.lcme.org].

MacQueen, K.M., McLellan, E., Kay, K. & Milstein, B. (1998) *Codebook Development for Team-based Qualitative Analysis* (Atlanta, GA, Centers for Disease Control).

Miller, G.E. (1990) The assessment of clinical skills/competence/performance, *Academic Medicine*, 65(Suppl.), pp. S63–S67.

Murphy, E., Dingwall, R., Greatbatch, D., Parker, S. & Watson, P. (1998) Qualitative research methods in health technology assessment: a review of the literature, *Health Technology Assessment*, 2(16).

Norman, G. (2002) Research in medical education: three decades of progress, *British Medical Journal*, 324, pp. 1560–1562.

Pressley, M. & McCormick, C.B. (1995) *Advanced Educational Psychology for Educators, Researchers, and Policymakers* (New York, HarperCollins College Publishers).

Probert, C.S., Cahill, D.J., McCann, G.L. & Ben-Shlomo, Y. (2003) Traditional finals and OSCEs in predicting consultant and self-reported clinical skills of PRHOs: a pilot study, *Medical Education*, 37, pp. 597–602.

Rubin, P. & Franchi-Christopher, D. (2002) New Edition of Tomorrow's Doctors. *Medical Teacher*, 24(4), pp. 368–369.

Scottish Deans Medical Curriculum Group (2002) The Scottish Doctor—Undergraduate Learning outcomes and their assessment: a foundation for competent and reflective practitioners [http://www.scottishdoctor.org].

Simpson, J.G., Furnace, J., Crosby, J., Cumming, A.D, *et al.* (2002) The Scottish doctor—learning outcomes for the medical undergraduate in Scotland: a foundation for competent and reflective practitioners, *Medical Teacher*, 24(2), pp. 136–143.

Smith, S.R. & Dollase, R. (1999) Planning, implementing and evaluating a competency-based curriculum, *Medical Teacher*, 21(1), pp. 15–22.

Smith, S.R. (1999) An Educational Blueprint for the Brown University School of Medicine, institutional publication, Brown University.

Snadden, D. (1999) AMEE Education Guide No. 11 (revised): Portfolio-based learning and assessment in medical education, *Medical Teacher*, 4, pp. 370–386.

Tamblyn, R.M., Klass, D.J., Schnabl, G.K. & Kopelow, M.L. (1991) Sources of unreliability and bias in standardized-patient rating, *Teaching and Learning in Medicine*, 3(2), pp. 74–85.

Van der Vleuten, C.P.M. (1996) The assessment of professional competence: developments, research and practical implications, *Advances in Health Sciences Education*, 1(1), pp. 41–67.

Van der Vleuten, C.P.M. (2000) *A* paradigm shift in education: how to proceed with assessment, oral presentation given at the 9th International Ottawa Conference on Medical Education, Cape Town, South Africa, 28 February–3 March [http://www.educ.unimaas.nl/ottawa/].

Van der Vleuten, C.P.M. & Swanson, D. (1990) Assessment of clinical skills with standardized patients: state of the art, *Teaching and Learning in Medicine*, 2, pp. 58–76.

Vu, N.V., Marcy, M.M., Colliver, J.A., Verhulst, S.J., Travis, T.A. & Barrows, H.S. (1992) Standardized (simulated) patients' accuracy in recording clinical performance checklist items, *Medical Education*, 16, pp. 99–104.

Winckel, C.P., Reznick, R., Cohen, R. & Taylor, B. (1994) Reliability and construct validity of a structured technical skills assessment form, *American Journal of Surgery*, 167, 423–427.

Wong, J.G.W.S. & Cheung, E.P.T. (2003) Ethics assessment in medical students, *Medical Teacher*, 25(1), pp. 5–8.