Workplace-based assessment as an educational tool

John Norcini & Vanessa Burch

Workplace-based assessment as an educational tool

John Norcini & Vanessa Burch

Institution/Corresponding Address:

Foundation for Advancement of International Medical Education and Research (FAIMER)
4th Floor, 3624 Market St
Philadelphia
PA 19104
USA

Tel: +1 215 823 2170

Fax: +1 215 386 2321

E mail: JNorcini@faimer.org

The authors

John J. Norcini, Ph.D. has been President and CEO of the Foundation for Advancement of International Medical Education and Research (FAIMER®) since May 2002. The Foundation's mission is to improve international medical education. FAIMER has a database of recognised medical schools around the world, an active research programme on international medical education and physician migration, and fellowship programmes designed for faculty from medical schools in developing countries.

For the 25 years before joining the Foundation, Dr. Norcini held a number of senior positions at the American Board of Internal Medicine. His principal academic interest is in the area of the assessment of physician performance. Dr. Norcini has published extensively, lectured and taught in many countries, and is on the editorial Boards of seven peer-reviewed journals in educational measurement and medical education.

Vanessa C. Burch, MBChB, PhD is Associate Professor of Medicine at the University of Cape Town, South Africa. She convenes the undergraduate medical degree programme in the Faculty of Health Sciences and is also actively involved in postgraduate education in the Faculty. Her main academic interests are in the assessment of clinical competence and innovative methods of medical education in resource-constrained educational environments typical of developing countries.

Workplace-based assessment as an educational tool

Contents

Abstract

There has been concern that trainees are seldom observed, assessed, and given feedback during their workplace-based education. This has led to an increasing interest in a variety of formative assessment methods that require observation and offer the opportunity for feedback.

The research literature on formative assessment and feedback suggests that it is a powerful means for changing the behaviour of trainees. Several methods for assessing it have been developed and there is preliminary evidence of their reliability and validity. A variety of factors enhance the efficacy of workplace-based assessment including the provision of feedback that is consistent with the needs of the learner and focused on important aspects of the performance. Faculty plays a critical role and successful implementation requires that they receive training.

The objective of this Guide is to review the literature on work-based assessment and the efficacy and prevalence of formative feedback. It describes the common formative assessment methods, characterises the nature of feedback, examines the effect of faculty development on its quality, and summarises the challenges still faced.

"A variety of factors enhance the efficacy of workplace-based assessment including the provision of feedback that is consistent with the needs of the learner and focused on important aspects of the performance."

Box 1

Take home messages

- The research literature on formative assessment and feedback suggests that it is a powerful means for changing the behaviour of learners.

- Several formative assessment methods have been developed for use in the workplace and there is preliminary data evidence of their reliability and validity.

- The efficacy of feedback is enhanced if it is consistent with the needs of the learner, focuses on important aspects of the performance, and has characteristics such as being timely and specific.

- Faculty development is critical to the quality and effectiveness of formative assessment.

- Strategies to encourage the participation of faculty are critical to the successful implementation of formative assessment.

Introduction

For just over two decades leading educationists, including medical educators, have highlighted the intimate relationship between learning and assessment. Indeed, in an educational context it is now argued that learning is the key purpose of assessment (van der Vleuten, 1996; Gronlund, 1998, Shepard, 2000). At the same time as this important connection was being stressed in the education literature, there were increasing concerns about the workplace-based training of doctors. A study (Day et al. 1990) in the United States documented that the vast majority of first-year trainees in internal medicine were not observed more than once by a faculty member in a patient encounter where they were taking a history or doing a physical examination. Without this observation, there was no opportunity for the assessment of basic clinical skills and, more importantly, the provision of feedback to improve performance.

As one step in encouraging the observation of performance by faculty, the American Board of Internal Medicine proposed the use of the mini-Clinical Evaluation Exercise (mini-CEX) (Norcini et al., 1995). In the mini-CEX, a faculty member observes a trainee as he/she interacts with a patient around a focused clinical task. Afterwards, the faculty member assesses the performance and provides the trainee feedback. It was expected that trainees would be assessed several times throughout the year of training with different faculty and in different clinical situations.

An advantage of the mini-CEX and other workplace-based methods is that they fulfil the three basic requirements for assessment techniques that facilitate learning (Frederiksen, 1984; Crooks, 1988; Swanson et al., 1995; Shepard, 2000): 1). The content of the training programme, the competencies expected as outcomes, and the assessment practices are aligned; 2) Trainee feedback is provided during and/or after assessment events; and 3) Assessment events are used strategically to steer trainee learning towards the desired outcomes. Over the past several years there has been growing interest in workplace-based assessment and additional methods have been (re)introduced to the setting of clinical training (National Health Service, 2007).

"An advantage of the mini-CEX and other workplace-based methods is that they fulfil the three basic requirements for assessment techniques that facilitate learning"

Previous publications have focused on the advantages and disadvantages of workplace-based methods from the perspective of assessment alone (Norcini, 2007). In this role, the methods are best thought of as analogous to classroom tests and they have much strength from this perspective. However, it is difficult to assure equivalence across institutions and the observations of faculty may be influenced by the stakes and their relationships with trainees. Consequently, their use faces challenges as national high stakes assessment devices.

Perhaps more importantly, workplace-based assessment can be instrumental in the provision of feedback to trainees to improve their performance and steer their learning towards desired outcomes. This guide focuses on the use of the methods for this purpose and it is divided into five sections. The first section briefly reviews the literature on the efficacy and prevalence of formative assessment and feedback. This is followed by a section that describes some of the more common methods of work-based assessment. The third section concentrates on feedback and it is explored from the perspective of the learner, its focus, and which characteristics make it effective in the context of formative assessment. Faculty play a key role in the successful implementation of formative assessment, so the fourth section describes strategies to encourage their participation and training to improve their performance. In the closing section of the guide we draw attention to the challenges faced by medical educators implementing formative assessment strategies in routine clinical teaching practice.

"Workplace-based assessment can be instrumental in the provision of feedback to trainees to improve their performance and steer their learning towards desired outcomes."

Efficacy and prevalence of formative assessment and feedback

Purpose of formative assessment and feedback

Formative assessment is not merely intended to assign grades to trainee performance at designated points in the curriculum; rather it is designed to be an ongoing part of the instructional process and to support and enhance learning (Shepard, 2000). Clearly, feedback is a core component of formative assessment (Sadler, 1989), central to learning, and at "the heart of medical education" (Branch & Paranjape, 2002). In fact, it is useful to consider feedback as part of an ongoing programme of assessment and instruction rather than a separate educational entity (Hattie & Timperley, 2007).

Feedback promotes student learning in three ways (Gipps, 1999; Shepard, 2000):

- It informs trainees of their progress or lack thereof;

- It advises trainees regarding observed learning needs and resources available to facilitate their learning; and

- It motivates trainees to engage in appropriate learning activities.

Efficacy of feedback

Given these presumed benefits, it is appropriate to ask whether there is a body of research supporting the efficacy of feedback in changing trainees' behaviour. Most compelling is a synthesis of information on classroom education by Hattie (1999) which included over 500 meta-analyses involving 1,800 studies and approximately 25 million students. He demonstrated that the typical effect size (ES) of schooling on overall student achievement is about 0.40 (i.e., it increases the mean on an achievement test by .4 of a standard deviation). Using this as a benchmark or "gold standard" on which to judge the various factors that affect performance, Hattie (1999) summarized the results of 12 meta-analyses that specifically included the influence of feedback. The feedback effect size was .79, which is certainly very powerful, and among the four biggest influences on achievement. Hattie (1999) also found considerable variability based on the type of feedback, with the largest effect being generated by the provision of information around a specific task.

Data to answer the question about the efficacy of feedback are much more limited in the domain of medical education but a recent meta-analysis by Veloski et al (2006) looked at its effect on clinical performance. Of the 41 studies meeting the criteria for inclusion, 74% demonstrated a positive effect for feedback alone. When combined with other educational interventions, feedback had a positive effect in 106 of the 132 (77%) studies reviewed.

A recent paper by Burch and colleagues reports on the impact of a formative assessment strategy implemented in a 4th year undergraduate medical clerkship programme (Burch et al., 2006). In this paper, students who engaged in an average of 6 directly observed clinical encounters during a 14-week clerkship reported that they more frequently undertook blinded patient encounters (Mcleod & Meagher, 2001) in which they did not consult the patient records before interviewing and examining the patient. Prior to implementing the formative assessment programme, students traditionally interviewed and examined patients only after consulting patient records. In addition they reported more frequent reading of topics relevant to patients clerked in the ward. While this paper only

"it is useful to consider feedback as part of an ongoing programme of assessment and instruction rather than a separate educational entity"

provides information on self-reported learning behaviour changes, it does suggest that formative assessment may have the potential to strategically direct student learning by reinforcing desirable learning behaviour (Gibbs, 1999).

A recent publication by Driessen and van der Vleuten (2000) supports the findings reported by Burch and colleagues. In their study they introduced a portfolio of learning assignments as an educational tool in a legal skills training programme comprising tutorials which were poorly attended and for which students did not adequately complete the required pre-tutorial work. The portfolio assignments, such as writing a legal contract or drafting a legislative document, were reviewed by peers and the tutor prior to being used as the teaching basis for subsequent skills training sessions. This educational intervention resulted in a twofold increase in time spent preparing for skills training sessions.

Prevalence of feedback

It is clear from these data that formative assessment and feedback have a powerful influence on trainee performance. However, there is a significant gap between what should be done and "on the ground" practice. Lack of assessment and feedback, based on observation of performance in the workplace, is one of the most serious deficiencies in current medical education practice (Holmboe, 2004; Kassebaum & Eaglen, 1999). Indeed, direct observation of trainee performance appears to be the exception rather than the rule.

In a survey of 97 United States medical schools, accredited between 1993 and 1998, it was found that structured, observed assessments of students' clinical abilities were done across clinical clerkships for only 7.4% to 23.1% of medical students (Kassebaum and Eaglen, 1999). A more recent survey of medical graduates found that during any given core clerkship, 17% to 39% of students were not observed performing a clinical examination (Association of American Medical Colleges, 2004). Likewise, Kogan and Hauer found that only 28% of Internal Medicine clerkships included an in-course formative assessment strategy involving observation of student performance in the workplace setting (Kogan & Hauer, 2006). Outside the US, Daelmans and colleagues (2004) reported that over a 6-month period, observation of trainee performance occurred in less than 35% of educational events in which observation and the provision of feedback could have taken place.

Unfortunately the situation is no better in postgraduate training programmes. In one study, 82% of residents reported that they engaged in only one directly observed clinical encounter in their first year of training; far fewer (32%) engaged in more than one encounter (Day et al, 1990). In another survey of postgraduate trainees 80% reported never or only infrequently receiving feedback based on directly observed performance (Isaacson et al, 1995).

Not only is assessment of directly observed performance infrequently done as part of routine educational practice, but the quality of feedback, when given, may be poor. Holmboe and colleagues evaluated the type of feedback given to residents after mini-CEX encounters and observed that while 61% of feedback sessions included a response from the trainee to the feedback, only 34% elicited any form of self-evaluation by the trainee. Of greatest concern, however, was the finding that only 8% of mini-CEX encounters translated into a plan of action (Holmboe et al., 2004). The paper by Holmboe and colleagues suggests that there are key reasons why clinician-educators fail to give trainees effective feedback (see Box 2).

"Lack of assessment and feedback, based on observation of performance in the workplace, is one of the most serious deficiencies in current medical education practice"

In addition to finding that trainee observation and feedback is infrequently given and often of limited value, it has also been noted that the faculties' assessment of trainee performance may be less than completely accurate. Noel and colleagues found that faculty failed to detect 68% of errors committed by postgraduate trainees when observing a videotape scripted to depict marginal competence (Noel et al., 1992). The use of checklists prompting faculty to look for specific skills increased error detection from 32% to 64%. It was, however, noted that this did not improve the accuracy of assessors. Approximately two thirds of faculty still scored the overall performance of marginal postgraduate trainees as satisfactory or superior. Similar observations attesting to the poor accuracy of faculty observations have been made elsewhere (Herberts et al, 1989; Kalet et al., 1992).

Based on the infrequency with which trainees are observed and problems with the quality of the feedback they receive, it is fair to ask whether observation of trainee performance is an outdated approach to medical training and assessment. The critical question, therefore, is whether clinical interviewing and examination skills are still relevant to clinical practice such that faculty should be trained to properly observe performance and provide effective, useful feedback.

Box 2

Key reasons why clinician-educators fail to give trainees effective feedback

- Current in-vivo assessment strategies such as the mini-CEX may be focusing on assessment of performance at the expense of providing adequate feedback;

- The scoring sheets currently used for in-vivo assessment events provide only limited space for recording comments thereby limiting feedback given;

- Clinician-educators do not fully appreciate the role of feedback as a fundamental clinical teaching tool;

- Clinician-educators may not be skilled in the process of providing high quality feedback.

Feedback in relation to history and physical examination

Despite major technological advances, the ability to competently interview and examine patients remains one of the mainstays of clinical practice (Holmboe, 2004). Data gathered over the past 30 years highlight the critical importance of these skills. In 1975 Hampton and colleagues demonstrated that a good medical history produced the final clinical diagnosis in 82% of 80 patients interviewed and examined. In only one of 80 cases did laboratory tests provide the final diagnosis not made by history or physical examination (Hampton et al., 1975).

Technological advances over the past two decades have not made the findings of this study irrelevant. In 1992 Peterson and colleagues showed that among 80 patients presenting for the first time to a primary care clinic, the patient's history provided the correct final diagnosis in 76% of cases (Peterson et al., 1992). Even more recently, an autopsy study of 400 cases showed that the combination of a history and physical examination produced the correct diagnosis in 70% of cases. Diagnostic imaging studies successfully indicated the correct diagnosis in only 35% of cases (Kirch & Schafii, 1996).

"Based on the infrequency with which trainees are observed and problems with the quality of the feedback they receive, it is fair to ask whether observation of trainee performance is an outdated approach to medical training and assessment."

Beyond diagnostic accuracy, physician-patient communication is a key component of health care. In a review of the literature, Beck et al. (2002) found that both verbal behaviors (e.g., empathy, reassurance and support) and nonverbal behaviors (e.g., nodding, forward lean) were positively associated with patient outcomes. Likewise, a study by Little et al. (2001) found that the patients of doctors who took a patient-centred approach were more satisfied, more enabled, had greater symptom relief, and had lower rates of referral.

The ability to competently interview a patient and perform a physical examination thus remains the cornerstone of clinical practice. The ability of faculty to accurately observe trainees performing these tasks and provide effective feedback is thus one of the most important aspects of medical training. Although methods such as standardised patients certainly provide complementary assessment and feedback information, they cannot replace the central role of observation by faculty.

Formative assessment methods

A number of assessment methods, suitable for providing feedback based on observation of trainee performance in the workplace, have been developed or regained prominence over the past decade. This section provides a brief description of the essential features of some of them including:-

- mini-clinical evaluation exercise (mini-CEX)

- clinical encounter cards (CEC)

- clinical work sampling (CWS)

- blinded patient encounters (BPE)

- direct observation of procedural skills (DOPS)

- case-based discussion (CbD)

- multisource feedback (MSF)

Mini-clinical evaluation exercise (mini-CEX)

As described above, the mini-CEX (Figure 1) is an assessment method developed in the United States (US) that is now in use in a number of institutions around the world. It requires trainees to engage in authentic workplace-based patient encounters while being observed by faculty members (Norcini et al., 1995). Trainees perform clinical tasks, such as taking a focused history or performing relevant aspects of the physical examination, after which they provide a summary of the patient encounter along with next steps (e.g., a clinical diagnosis and a management plan).

These encounters can take place in a variety of workplace settings including inpatient, outpatient, and emergency departments. Patients presenting for the first time as well as those returning for follow up visits are suitable encounters for the mini-CEX. Not surprisingly, the method lends itself to a wide range of clinical problems including: (1) presenting complaints such as chest pain, shortness of breath, abdominal pain, cough, dizziness, low back pain; or (2) clinical problems such as arthritis, chronic obstructive airways disease, angina, hypertension and diabetes mellitus (Norcini et al., 2003).

"The ability of faculty to accurately observe trainees performing these (history-taking and physical examination) tasks and provide effective feedback is thus one of the most important aspects of medical training."

Figure 1. Mini-clinical evaluation exercise form.

Please refer to www.hcat.nhs.uk for guidance on this form and details of expected competencies for F1

Mini-Clinical Evaluation Exercise (CEX) - F1 Version

Please complete the questions using a cross: ☒   Please use black ink and CAPITAL LETTERS

Doctor's Surname: ☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐

Forename: ☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐

GMC Number: ☐☐☐☐☐☐☐   **GMC NUMBER MUST BE COMPLETED**

Clinical setting: A&E ☐   OPD ☐   In-patient ☐   Acute Admission ☐   GP Surgery ☐

Clinical problem category: Airway/Breathing ☐   CVS/Circulation ☐   Gastro ☐   Neuro ☐   Pain ☐   Psych/Behav ☐   Other ☐ _____

New or FU: New ☐   FU ☐   Focus of clinical encounter: History ☐   Diagnosis ☐   Management ☐   Explanation ☐

Number of times patient seen before by trainee: 0 ☐   1-4 ☐   5-9 ☐   >10 ☐   Complexity of case: Low ☐   Average ☐   High ☐

Assessor's position: Consultant ☐   GP ☐   SpR ☐   SASG ☐   SHO ☐   Other ☐ _____

Number of previous mini-CEXs observed by assessor with <u>any</u> trainee: 0 ☐   1 ☐   2 ☐   3 ☐   4 ☐   5-9 ☐   >9 ☐

Please grade the following areas using the scale below:	Below expectations for F1 completion		Borderline for F1 completion	Meets expectations for F1 completion	Above expectations for F1 completion		U/C*
1. History Taking	☐	☐	☐	☐	☐	☐	☐
2. Physical Examination Skills	☐	☐	☐	☐	☐	☐	☐
3. Communication Skills	☐	☐	☐	☐	☐	☐	☐
4. Clinical Judgement	☐	☐	☐	☐	☐	☐	☐
5. Professionalism	☐	☐	☐	☐	☐	☐	☐
6. Organisation/Efficiency	☐	☐	☐	☐	☐	☐	☐
7. Overall clinical care	☐	☐	☐	☐	☐	☐	☐

*U/C Please mark this if you have not observed the behaviour and therefore feel unable to comment.

Anything especially good? | **Suggestions for development**

Agreed action:

Have you had training in the use of this assessment tool?: ☐ Face-to-Face   ☐ Have Read Guidelines   ☐ Web/CD rom

Assessor's Signature: _____   Date (mm/yy): M M / Y Y   Time taken for observation: (in minutes) ☐☐

Assessor's Surname: ☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐   Time taken for feedback: (in minutes) ☐☐

Assessor's registration number: ☐☐☐☐☐☐☐☐☐☐☐

Please note: Failure of return of all completed forms to your administrator is a probity issue
Acknowledgements: Adapted with permission from American Board of Internal Medicine

Source: www.hcat.nhs.uk

In the original work, each aspect of the clinical encounter is scored by a faculty member using a 9-point rating scale where 1-3 is unsatisfactory, 4-6 is satisfactory and 7-9 is superior. The parameters evaluated include: interviewing skill, physical examination, professionalism, clinical judgement, counselling, organization and efficiency, and overall competence. Different scales and different parameters have been used successfully in other settings (e.g., National Health Service).

The core purpose of the assessment method is to provide structured feedback based on observed performance. Each patient encounter takes roughly 15 minutes followed by 5-10 minutes of feedback. Trainees are expected to be evaluated several times with different patients and by different faculty members during their training period (e.g., the Foundation Programme requires 6 assessments per year).

This assessment tool has been shown to be a reliable way of assessing postgraduate trainee performance provided there is sufficient sampling. Roughly 4 encounters are sufficient to achieve a 95% confidence interval of less than 1 (on the 9-point scale) and approximately 12-14 are required for a reliability coefficient of 0.8 (Norcini et al. 1995; Norcini et al., 2003; Holmboe et al., 2003).

In addition to the postgraduate setting, the mini-CEX has been successfully implemented in undergraduate medical training programmes (Hauer, 2000; Kogan et al., 2003; Kogan & Hauer, 2006). In this context, the period of observation and feedback is often longer, ranging from 30-45 minutes (Hauer, 2000; Kogan et al., 2002).

There is a growing body of evidence supporting the validity of the mini-CEX. Kogan et al (2002 and 2003) found that mini-CEX performance was correlated with other assessments collected as part of undergraduate training. Faculty ratings of videotapes of student-standardised patient encounters, using the mini-CEX forms, were correlated with the checklist scores and standardised patient ratings of communication skills (Boulet et al, 2002). In postgraduate training, mini-CEX performance was correlated with a written in-training examination and routine faculty ratings (Durning et al, 2002). Holmboe et al (2004) found that, using the mini-CEX form, they could differentiate amongst videos, scripted to represent different levels of ability. Finally, Hatala et al (2006) found that mini-CEX scores were correlated with the results of a Royal College oral examination.

"Mini-CEX requires trainees to engage in authentic workplace-based patient encounters while being observed by faculty members."

Clinical encounter cards (CEC)

The CEC system, developed at McMaster University in Canada (Hatala & Norman, 1999) and subsequently implemented in other centres (Paukert et al., 2002), is similar to the mini-CEX. The basic purpose of this assessment strategy is also to score trainee performance based on direct observation of a patient encounter. The encounter card system scores the following dimensions of observed clinical practice: history-taking, physical examination, professional behaviour, technical skill, case presentation, problem formulation (diagnosis) and problem solving (therapy). Each dimension is scored using a 6-point rating scale describing performance as 1) unsatisfactory, 2) below the expected level of student performance, 3) at the expected level of student performance, 4) above the expected level of student performance, 5) outstanding student performance, and 6) performance at the level of a medical graduate.

In addition to capturing the quality of the performance, the 4 x 6 inch score cards also provide space for assessors to record the feedback given to the trainee at the end of the encounter. This system has been shown to be a feasible, valid, and reliable measure of clinical competence, provided that a sufficient number of

"The encounter card system scores the following dimensions of observed clinical practice: history-taking, physical examination, professional behaviour, technical skill, case presentation, problem formulation (diagnosis) and problem solving (therapy)."

encounters (approximately 8 encounters for a reliability coefficient of 0.8 or more) are collected (Hatala & Norman, 1999). Moreover, introduction of the system was found to increase student satisfaction with the feedback process (Paukert et al., 2002) and to have modest correlations with other forms of assessment (Richards et al, 2007).

Clinical work sampling (CWS)

This assessment method, developed in Canada, is also based on direct observation of clinical performance in the workplace (Turnbull et al., 2000). The method requires collection of data concerning specific patient encounters for a number of different domains either at the time of admission (admission rating form) or during the hospital stay (ward rating form). Both of these forms are completed by faculty members directly observing trainee performance. The domains assessed by faculty include: communication skills, physical examination skills, diagnostic acumen, consultation skills, management skills, interpersonal behaviour, continued learning skills and health advocacy skills. Not all skills are evaluated on each occasion.

Trainees are also assessed by ward nursing staff (using the multidisciplinary team rating form) and the patients (using the patient rating form) who are in the care of the trainees. These rating forms, also completed on the basis of directly observed behaviour, require a global assessment and ratings of the following domains: therapeutic strategies, communications skills, consultation with other health care professionals, management of resources, discharge planning, interpersonal relations, collaboration skills, and health advocacy skills and professionalism.

All rating forms use a 5-point rating scale ranging from unsatisfactory to excellent performance. This assessment method has also been shown to be valid and reliable provided a sufficient number (approximately 7 encounters for a reliability coefficient of 0.7) of encounters are observed (Turnbull et al., 2000).

A later study found that the CWS strategy could be adapted to radiology residency using a handheld computerized device (Finlay et al, 2006). Compliance with voluntary participation was not as great as expected but this evaluation format included the opportunity to discuss performance at the time of data entry, rather than at the end of rotation. The investigators found the method less useful for summative purposes although the sample size was small (N=14).

Blinded patient encounters

This formative assessment method is based on the same principle as the three assessment methods already mentioned. It is unique, however, in that it forms part of undergraduate bedside teaching sessions. (Burch et al., 2006). Students, in groups of 4-5, participate in a bedside tutorial. It starts with a period of direct observation in which one of the students in the group is observed performing a focused interview or physical examination as instructed by the clinician educator conducting the teaching session. Thereafter the student is expected to provide a diagnosis, including a differential diagnosis, based on the clinical findings.

The patient is unknown to the student, hence the term "blinded" patient encounter (McLeod & Meagher, 2001). This type of patient encounter has the advantage of safely allowing the trainee to practice information gathering, hypothesis generation, and problem solving without access to the workup by more senior doctors.

"Clinical work sampling requires collection of data concerning specific patient encounters for a number of different domains either at the time of admission (admission rating form) or during the hospital stay (ward rating form)."

"It (blinded patient encounters) is unique, however, in that it forms part of undergraduate bedside teaching sessions."

Figure 2. Directly observed procedural skills form.

Source: www.hcat.nhs.uk

After the presentation, the session focuses on demonstrating the important clinical features of the case as well as discussing various issues, for example appropriate investigation and treatment relevant to the patient's presenting clinical problem. It concludes with a feedback session in which the student receives personal private advice about his/her performance.

Feedback is provided using a 9-point rating scale for assessment of clinical interviewing and examination skills as well as clinical reasoning skills. The rating scale ranges from 1-3 for poor performance, 4-6 for adequate performance and 7-9 for good performance. Space is provided on the score sheet to add other written comments. Students keep the score sheets which are only used for feedback purposes.

Direct observation of procedural skills (DOPS)

This assessment method (Figure 2), developed in the UK, focuses on evaluating the procedural skills of postgraduate trainees by observing them in the workplace setting (Wragg et al., 2003). Just as in CWS and the encounter card assessment systems, trainees' performance is scored using a 6-point rating scale where 1-2 is below the expected level of competency, 3 reflects a borderline level of competency, 4 meets the expected level of competency and 5-6 are above the expected level of competency. The assessment procedure is generally expected to require 15 minutes of observation time and 5 minutes dedicated to feedback.

Trainees are provided with a list of commonly performed procedures for which they are expected to demonstrate competence such as endotracheal intubation, nasogastric tube insertion, administration of intravenous medication, venepuncture, peripheral venous cannulation and arterial blood sampling. They are assessed by multiple clinicians on multiple occasions throughout the training period.

This method of procedural skills assessment is not limited to postgraduate training programmes. Paukert and colleagues have included basic surgical skills to be mastered by undergraduate students in their clinical encounter card system (Paukert et al., 2002).

Although DOPS is similar to procedural skills log books, the purpose and nature of these methods differ significantly. The recording of procedures is common to both of them, but log books are usually designed to ensure that trainees have simply performed the minimum number required to be considered competent. The provision of structured feedback based on observation of a performance is not necessarily part of the log book process. Moreover, the procedure is not necessarily performed under direct observation and little feedback, if any, is expected to be given. In contrast, DOPS ensures that trainees are given specific feedback based on direct observation so as to improve their procedural skills.

Case-based discussion (CbD)

This assessment method is an anglicised version of Chart-Stimulated Recall (CSR) developed for use by the American Board of Emergency Medicine (Maatsch et al., 1983). It is currently part of the Foundation Programme implemented for postgraduate training in the UK National Health Service (Figure 3). In CbD, the trainee selects two case records of patients in which they had made notes and presents them to an assessor. The assessor selects one of the two for discussion and explores one or more aspects of the case, including: clinical assessment, investigation and referral of the patient, treatment, follow-up and future planning, and professionalism (Figure 3). Since the case record is available at the time of assessment, medical record keeping can also be assessed by the examiner.

"DOPS ensures that trainees are given specific feedback based on direct observation so as to improve their procedural skills."

Figure 3. Case-based assessment form.

Source: www.mmc.nhs.uk

This type of performance assessment focuses on evaluating the clinical reasoning of trainees so as to understand the rationale behind decisions made in authentic clinical practice. As with other assessment methods described, each encounter is expected to last no more than 20 minutes, including 5 minutes of feedback. Trainees are expected to engage in multiple encounters with multiple different examiners during the training period.

There are several studies supporting the validity of this measure. Maatsch et al (1983) collected several assessments for a group of practicing doctors eligible for recertification in Emergency Medicine. They found that CbD correlated with a number of the other measures, including chart audit. The score distribution and pass-fail results were consistent with scores on initial certification, ten years earlier. As importantly, CbD was considered the most valid of the measures by the practicing doctors participating in the study.

A study by Norman et al. (1989) compared a volunteer group of doctors to those referred for practice difficulties. CbD was highly correlated with a standardised patient examination and with an oral examination. More importantly, it was able to separate the volunteer group from the doctors who were referred. Likewise, Solomon et al (1990) collected data from several different assessments on practicing doctors eligible for recertification. CbD was correlated with the oral examination as well as written and oral exams administered 10 years earlier.

Multisource feedback (MSF)

More commonly referred to as 360-degree assessment, this method represents a systematic collection of performance data and feedback for an individual trainee, using structured questionnaires completed by a number of stakeholders. The assessments are all based on directly observed behaviour (Wragg et al., 2003) but they differ from the methods presented above in that they reflect routine performance, rather than performance during a specific patient encounter.

Although there are a number of different ways of conducting this form of assessment, the mini-peer assessment tool (mini-PAT) that has been selected for use in the Foundation Programme in the UK is a good example. Trainees nominate 8 assessors including senior consultants, junior specialists, nurses and allied health service professionals. Each of the nominated assessors receives a structured questionnaire (Figure 4) which is completed and returned to a central location for processing. Trainees also complete self-assessments, using the same questionnaires, and submit these for processing. The categories of assessment include: good clinical care, maintaining good clinical practice, teaching and training, relationships with patients, working with colleagues and an overall assessment.

The questionnaires are collated and individual feedback is prepared for trainees. Data are provided in a graphic form which depicts the mean ratings of the assessors and the national mean rating. All comments are included verbatim, but they remain anonymous. Trainees review this feedback with their supervisor and together work on developing an action plan. This process is repeated twice yearly during the training period.

This method is widely used in industry and business, but has also been found to be useful in medicine. Applied to practicing doctors, it was able to distinguish certified from non-certified internists and the results were associated with performance on a written examination (Ramsey et al., 1989; Wenrich et al., 1993). In a follow-up study, two subscales were identified—one focused on technical/cognitive skills and the other focused on professionalism (Ramsey et al., 1993). Written examination performance was correlated with the former but not the latter.

Multisource feedback has been applied to postgraduate trainees as well as practicing doctors. The Sheffield Peer Review Assessment Tool, which is the full scale version of mini-PAT as shown in Figure 4 was studied with paediatricians and found to be feasible and reliable (Archer et al., 2005). It also separated doctors by grade and tended to be insensitive to potential biasing factors such as the length of the working relationship. Whitehouse et al. (2002) also applied multisource feedback to postgraduate trainees with reasonable results.

Finally, this form of assessment has also been used successfully with medical students (Arnold et al., 1981; Small et al., 1993). Both positive and negative reports from peers have influenced academic actions.

Overall, reasonably reliable results can be achieved with the assessments of 8 to 12 peers.

Nature of the feedback

For the purpose of this discussion, feedback can be conceptualised as "information provided by an agent (teacher, peer, self, etc.) regarding aspects of one's performance or understanding" (Hattie & Timperley, 2007). This information can be used by the learner to "confirm, add to, overwrite, tune or restructure information in memory, whether that information is domain knowledge, meta-cognitive knowledge, belief about self and tasks or cognitive tactics and strategies" (Winnie & Butler, 1994). The main purpose of feedback is, therefore, to reduce the discrepancy between current practices or understandings and desired practices or understandings (Hattie & Timperley, 2007).

Perspective of the learner

In order for feedback to fulfil this purpose, it needs to address three fundamental questions for the learner:

• Where am I going?

• How am I going?

• Where to next?

To address the first question, it is critical that there be clearly defined learning goals. If the goals are not clearly articulated then "the gap between current learning and intended learning is unlikely to be sufficiently clear for students to see a need to reduce it " (Hattie & Timperley, 2007). Goals can be wide ranging and variable, but without them students are less likely to engage in properly directed action, persist at tasks in the face of difficulties, or resume the task if disrupted (Bargh et al., 2001). The existence of goals is also more likely to lead students to seek and receive feedback, especially if they have a shared commitment to achieving them (Locke & Latham, 1990). So, medical trainees need to have a clear understanding of desired practice or competence in order to seek feedback and stay focused on the task of achieving competence in the domain of interest.

The second question focuses on the provision of concrete information, derived from an assessment of the performance, relative to a task or goal. To do so well requires criteria that provide clear indicators of whether the task has been completed properly. The answer to this question addresses the traditional, restricted definition of feedback. Nonetheless, it is critical to the provision of effective feedback. Ironically, it is precisely this aspect of feedback which is usually poorly done. Clinician-educators are often reluctant to provide honest feedback, particularly in the face of poor performance.

Figure 4. Mini-peer assessment questionnaire

mini-PAT (Peer Assessment Tool) - F1 Version

Please complete the questions using a cross: ☒   Please use black ink and CAPITAL LETTERS

Doctor's   Surname

Forename

GMC Number:

How do you rate this Doctor in their:

		Below expectations for F1 completion	Borderline for F1 completion	Meets expectations for F1 completion	Above expectations for F1 completion		U/C*
		1	2	3	4	5	6	

Good Clinical Care

		1	2	3	4	5	6	U/C*
1	Ability to diagnose patient problems	☐	☐	☐	☐	☐	☐	☐
2	Ability to formulate appropriate management plans	☐	☐	☐	☐	☐	☐	☐
3	Awareness of their own limitations	☐	☐	☐	☐	☐	☐	☐
4	Ability to respond to psychosocial aspects of illness	☐	☐	☐	☐	☐	☐	☐
5	Appropriate utilisation of resources e.g. ordering investigations	☐	☐	☐	☐	☐	☐	☐

Maintaining good medical practice

6	Ability to manage time effectively / prioritise	☐	☐	☐	☐	☐	☐	☐
7	Technical skills (appropriate to current practice)	☐	☐	☐	☐	☐	☐	☐

Teaching and Training, Appraising and Assessing

8	Willingness and effectiveness when teaching/training colleagues	☐	☐	☐	☐	☐	☐	☐

Relationship with Patients

9	Communication with patients	☐	☐	☐	☐	☐	☐	☐
10	Communication with carers and/or family	☐	☐	☐	☐	☐	☐	☐
11	Respect for patients and their right to confidentiality	☐	☐	☐	☐	☐	☐	☐

Working with colleagues

12	Verbal communication with colleagues	☐	☐	☐	☐	☐	☐	☐
13	Written communication with colleagues	☐	☐	☐	☐	☐	☐	☐
14	Ability to recognise and value the contribution of others	☐	☐	☐	☐	☐	☐	☐
15	Accessibility/Reliability	☐	☐	☐	☐	☐	☐	☐
16	Overall, how do you rate this doctor compared to a doctor ready to complete F1 training?	☐	☐	☐	☐	☐	☐	☐

Do you have any concerns about this doctor's probity or health?   ☐ Yes   ☐ No

If yes please state your concerns:

*U/C Please mark this if you have not observed the behaviour and therefore feel unable to comment.   6927534062

Anything especially good?

Please describe any behaviour that has raised concerns or should be a particular focus for development:

Please continue your comments on a separate sheet if required

Your Gender: ☐ Male   ☐ Female

Your ethnic group:
- ☐ British
- ☐ Irish
- ☐ Other White Background
- ☐ Caribbean
- ☐ African
- ☐ Any other Black background
- ☐ Indian
- ☐ Pakistani
- ☐ Bangladeshi
- ☐ Other Asian Background
- ☐ White and Black Caribbean
- ☐ White and Black African
- ☐ White and Asian
- ☐ Any other mixed background
- ☐ Chinese
- ☐ Any other ethnic group

Which environment have you primarily observed the doctor in?
(Please choose one answer only)
- ☐ Inpatients
- ☐ Outpatients
- ☐ Both In and Out-patients
- ☐ A&E/Admissions
- ☐ Community Speciality
- ☐ Laboratory/Research
- ☐ Intensive Care
- ☐ Theatre
- ☐ General Practice
- ☐ Other (Please specify)

Your position:
- ☐ Consultant   ☐ SASG   ☐ SpR   ☐ Foundation/PRHO
- ☐ Nurse   ☐ SHO   ☐ Allied Health Professional
- ☐ GP
- ☐ Other (Please specify)

If you are a Nurse or AHP how long have you been qualified?:   ___ years

Length of working relationship: ___ months

What training have you had in the use of this assessment tool?:   ☐ Face-to-Face   ☐ Have Read Guidelines   ☐ Web/CD rom

How long has it taken you to complete this form (in minutes)?:   ___

Your Signature: ___   Date: __ / __ / __

Your Surname:

Your GMC Number: (Doctors only)

Source: www.mmc.nhs.uk

Having a set of clearly defined criteria makes it somewhat easier to provide guidance based strictly on observed performance, rather than interpretations of the trainee's intentions.

The final important question from the perspective of the trainee is what actions need to be taken in order to close the gap between actual performance and desired performance. Trainees need an action plan—specific information about how to proceed in order to achieve desired learning outcomes. As indicated previously, without honest feedback regarding actual performance, trainees are unlikely to seek advice about how to proceed in order to close the learning gap.

The interrelatedness of these questions becomes apparent when attempting to address this final question. Indeed, without clearly defined learning outcomes, including criteria which make achievement of the learning goals explicit, and honest feedback about observed performance, planning aimed at improving performance will not take place. Closing the gap between where trainees are and where they need to be is both the purpose of feedback and the source of its influence (Sadler, 1989).

Focus of feedback

How effectively feedback addresses the three questions for learners is dependent in part on what aspects of the performance are addressed. Specifically, there are four foci for feedback (Hattie & Timperley, 2007):

- Feedback about the task;

- Feedback about the process of the task;

- Feedback about self-regulation; and

- Feedback about the self as a person.

The most basic focus of feedback addresses the quality of the task performed. Using well defined criteria, trainees are given specific information about whether they achieved the required level of performance. This type of feedback is easiest to give, and is consequently the most frequently provided. It is most helpful when it concentrates on the performance, rather than the knowledge required for the task. The latter is best dealt with by providing direct instruction and it is not regarded as feedback (Hattie & Timperley, 2007).

One of the limitations of providing feedback focused only on the task is that it is necessarily context-specific or task-specific. Consequently, it does not generalise readily to other tasks (Thompson, 1998). On the other hand, providing feedback that focuses on the process can be of more value because it encourages a deeper appreciation of the performance. This involves giving feedback that enhances an understanding of relationships (i.e. the construction of meaning), cognitive processes, and transfer to different or novel situations (Marton et al., 1993). This focus for feedback is also more likely to promote deep learning (Balzer, 1989).

A major component of this type of feedback is the provision of strategies for error detection and correction, in other words developing the trainee's ability to provide self-feedback (Hattie & Timperley, 2007). Feedback about the process underlying the task can also serve as a cueing mechanism leading to more effective information search strategies. Cueing is most useful when it assists trainees in detecting faulty hypotheses and provides direction for further searching and strategising (Harackiewicz, 1979).

"The most basic focus of feedback addresses the quality of the task performed. Using well defined criteria, trainees are given specific information about whether they achieved the required level of performance."

Feedback that focuses on self-regulation addresses the interplay between commitment, control, and confidence. It concentrates on the way trainees monitor, direct, and regulate their actions relative to the learning goal. It implies a measure of autonomy, self-control, self-direction, and self-discipline (Hattie & Timperley, 2007). Effective learners are able to generate internal feedback and cognitive routines while engaged in a task (Butler & Winnie, 1995).

Students who are able to self-appraise and self-manage are able to seek and receive feedback from others. At the other end of the spectrum are less effective learners who, having minimal self-regulation strategies, are more dependent on external factors, such as teachers, to provide feedback. For these learners, feedback is more effective if it directs attention back to the task and enhances feelings of self-efficacy such that trainees are likely to invest more time and become more committed to mastering the task (Kluger & DeNisi, 1996).

Trainees' attributions of success and failure can have more impact than actual success or failure. Feelings of self-efficacy can be adversely affected if students are unable to relate feedback to the cause of their poor performance. In other words, feedback that does not specify the grounds on which students have achieved success or not, is likely to engender personal uncertainties and may ultimately lead to poorer performance (Thompson, 1998). On the other hand, feedback that attributes performance to effort or ability is likely to increase engagement and task performance (Craven et al., 1991). Thus, when giving feedback it is critical that the assessor clearly directs the feedback to observed performance, while being aware of the impact feedback has on the self-efficacy of the trainee.

The final focus of feedback is discussed not because of its educational value but rather because it often has adverse consequences. This feedback is typically concentrated on the personal attributes of the trainee and seldom contains task-related information, strategies to improve commitment to the task, or a better understanding of self or the task itself (Hattie & Timperley, 2007). This focus for feedback is generally not effective, its impact is unpredictable, and it can have an adverse effect on learning. This is particularly true of negative feedback directed at a personal level.

"When giving feedback it is critical that the assessor clearly directs the feedback to observed performance, while being aware of the impact feedback has on the self-efficacy of the trainee."

Characteristics of effective feedback in the context of formative assessment

Formative assessment strategies are thought to best prompt change when they are integral to the learning process, performance assessment criteria are clearly articulated, feedback is provided immediately after the assessment event, and trainees engage in multiple assessment opportunities (Crooks, 1988; Gibbs & Simpson, 2004). In addition to these features, Ende suggested that specific conditions could make feedback more conducive to learning (Ende, 1983) as described in Box 3.

Box 3

Specific conditions to make feedback more conducive to learning (Ende, 1983)

- Set an appropriate time and place for feedback

- Provide feedback regarding specific behaviours, not general performance

- Give feedback on decisions and actions, not one's interpretation of the trainee's motives or intentions

- Give feedback in small digestible quantities and

- Use language that is non-evaluative and non-judgemental.

In addition to the strategies suggested by Ende, it has also been suggested that the efficacy of feedback may be further improved by promoting trainee "ownership" of feedback (Holmboe et al., 2004). Strategies to achieve this include:

- Encouraging trainees to engage in a process of self-assessment prior to receiving external feedback;

- Permitting trainees to respond to feedback; and

- Ensuring that feedback translates into a plan of action for the trainee.

Based on a large qualitative study, including 83 academics involved in education, Hewson and Little (1998) validated many of these literature-based recommendations. They developed a useful list of bipolar descriptors outlining feedback techniques to be adopted and avoided (Box 4).

Box 4

Feedback techniques to be avoided and adopted (taken from Hewson & Little, 1998)

Feedback techniques to be avoided	Feedback techniques to be adopted
Creating a disrespectful, unfriendly, closed, threatening climate	Creating a respectful, open minded, non-threatening climate
Not eliciting thoughts or feelings before giving feedback	Eliciting thoughts and feelings before giving feedback
Being judgemental	Being non-judgemental
Focusing on personality	Focusing on behaviours
Basing feedback on hearsay	Basing feedback on observed facts
Basing feedback on generalizations	Basing feedback on specifics
Giving too much/too little feedback	Giving the right amount of feedback
Not suggesting ideas for improvement	Suggesting ideas for improvement
Basing feedback on unknown, non-negotiated goals	Basing feedback on well-defined, negotiated goals

As already mentioned, formulating an action plan at the end of a feedback session is critical to the success of formative assessment. If a plan addressing the deficiencies is not formulated, it results in failure to close the "learning loop" and correct the identified problems (Holmboe et al., 2004). Indeed, formulation of an action plan may constitute the most critical step in providing feedback.

Beyond these actions, it is becoming increasingly recognised that ongoing coaching or mentoring improves the efficacy of feedback. This is particularly true of 360-degree feedback strategies (Luthans & Peterson, 2004). Current literature in the business world reports that the role of the workplace managers has been reconceptualised such that they are seen to be facilitators of learning, creativity, and innovation rather than directors or controllers of activity. Furthermore, learning leaders or managers should foster interconnections between people and systems so as to create collective learning networks (Walker, 2001). While this research has not been replicated in the medical workplace setting, the emerging success of these strategies in business suggests that similar methods merit further consideration in clinical training settings.

Faculty development

Faculty participation

From the preceding discussion it is clear that there is a need to increase the frequency of observation of trainee performance in order to provide feedback aimed at improving the quality of the services they later render in clinical practice. To this end a number of strategies have recently been implemented, but the studies of their efficacy are limited in number and they report variable success.

Holmboe and colleagues examined the impact of a scoring sheet specifically designed to remind faculty both of the dimensions of feedback and that its main purpose is to provide trainees with information about their performance aimed at improving it (Holmboe et al., 2001). In the study, the faculty control group did not receive any instruction regarding the use of the score sheet, while the intervention group received 20 minutes of instruction at the start of the clinical rotation. This information session outlined the characteristics of effective feedback and stressed the importance of direct observation of trainees to evaluate clinical competence. Results of the study indicated that while the intervention group did not provide more frequent feedback, their trainees were more satisfied with the quality of feedback they received.

Two recent studies in the Netherlands have produced similar findings. In one of the studies an undergraduate surgical clerkship was restructured in an attempt to increase the observation of trainee performance and the provision of feedback by senior faculty members (van der Hem-Stokroos et al., 2004). Restructuring of the clerkship included the introduction of a log book, a form documenting observation of skill performance, and individual appraisal by senior staff. Faculty was informed of the changes but they were not given formal instruction in trainee observation and how to provide feedback. The results indicated no significant increase in trainee observation or the provision of feedback. The authors suggest that the lack of impact of the intervention may be partly attributed to the limited input received by faculty involved in the study, particularly limited involvement in the process of restructuring the clerkship.

In the other study, Daelmans and colleagues (Daelmans et al., 2005) introduced in-training assessment in an undergraduate medical clerkship programme. Senior clinical staff was informed about the introduction at a meeting held at the beginning of the clerkship. They also received a letter outlining the in-training assessment programme. The findings indicated that despite implementing this new programme, students were not more frequently observed performing clinical interviews and examinations in the workplace. In their discussion of the results they suggest that observation and feedback regarding student performance may have been improved if faculty members had been more frequently reminded of the programme, for example daily meetings could have been used to alert faculty to the importance and potential educational value of the programme.

In contrast to these studies, Turnbull and colleagues (Turnbull et al., 2000) describe a strategy using clinical work sampling in which students received feedback based on directly observed patient encounters an average of eight times during a 4-week clerkship rotation. In this study, faculty members observing students in the workplace attended a 2-hour workshop outlining the assessment and feedback strategy. In addition, they received monthly communications reminding them of the project. Students were also oriented to the project before it started, and met with the research associate on a weekly basis during the clerkship rotation. Results indicated that the ongoing collection of performance data was feasible.

"There is a need to increase the frequency of observation of trainee performance in order to provide feedback aimed at improving the quality of the services they later render in clinical practice."

In another study using the clinical encounter card system, students engaged in a directly observed assessment event an average of 35 times during a 12-week surgery clerkship (Paukert et al., 2002). As in the other study, evaluators involved in the project were briefed about the project in a number of short 15-minute meetings outlining the purpose and importance of the intervention implemented. These information sessions formed part of other meetings routinely held in the department, for example morbidity and mortality meetings. At each of these information sessions, faculty were asked to raise any issues or concerns they had regarding the project. They also received a letter explaining the assessment and feedback system prior to implementation. At the end of the clerkship, students were more satisfied with the feedback they received.

Based on these studies it is clear that a number of strategies need to be employed to successfully implement an assessment process in which trainees receive feedback based on directly observed performance in the workplace. First, it is apparent that involvement of faculty in planning an in-course formative assessment strategy is likely to enhance their engagement in the process. Second, faculty need to be thoroughly briefed about the purpose and process of the observation and feedback strategy implemented. Third, students need to be properly informed about the purpose and format of the assessment method used. In particular, it is critical that the potential learning benefits of the system are emphasized rather than the assessment aspects of the methods being used. Finally, faculty and students need to be regularly reminded of the benefit of formative assessment and the importance of keeping the assessment strategy active in the workplace.

Faculty training

While successfully implementing a formative assessment strategy in the workplace is an achievement in its own right, it is important to ensure that the quality of the observations made by attending faculty are accurate and that the feedback received by students is effective. As was highlighted earlier, faculty observations of student performance may not be sufficiently accurate to identify errors in student performance. While the use of checklists has been shown to improve the ability of assessors to detect errors in performance (Noel et al, 1992), they have not been shown to improve the overall accuracy of assessors. This is an issue that requires further research; effective strategies to address this problem clearly need to be found.

While the accuracy of examiners remains an issue needing further work, the stringency of examiners can be improved with training. A recent paper by Boulet and colleagues examined the stringency of examiners using the mini-CEX to evaluate directly observed trainee performance (Boulet, et al., 2002). They reported significant variability among the examiners even when they were observing the same event. Holmboe and colleagues have shown that assessor training can address this issue. In their paper, study participants engaged in a one-day video-based training session aimed at reducing variability among faculty when providing assessments and feedback on observed performance. Participants engaged in performance dimension training and frame-of-reference training (Holmboe et al., 2004). The former was accomplished by getting faculty to discuss and define key components of competence for specific clinical skills and develop criteria for satisfactory performance. The latter was addressed by giving individual faculty members the opportunity to score real-time trainee performance using standardised patients and standardised trainees. While one faculty member scored the performance of the trainee and provided feedback, other faculty members scored the trainee's performance by watching the interview and examination on a video monitor. The encounter ended with a group discussion

"Faculty observations of student performance may not be sufficiently accurate to identify errors in student performance"

of how each member of the group rated the performance and reasons for the scores allocated. Finally the facilitator described what type of trainee performance the case scenario was scripted to depict.

Eight months after this faculty development effort, a set of video recordings of scripted patient encounters were again used to compare the performance of trained faculty as compared to a cohort of untrained faculty. Trained faculty were more stringent than untrained faculty members and they also reported feeling more comfortable providing trainee feedback. This study is one of the first demonstrating the beneficial impact of faculty training for the purpose of scoring performance with the intention of providing trainee feedback.

Challenges

In this closing section of the guide we wish to highlight areas where further work is needed to address some pivotal questions regarding workplace-based formative assessment and feedback. First and foremost, we need to develop strategies that will ensure successful and sustainable implementation of formative assessment in the workplace. Most of what has been done to date has been research-based, short term projects. We need studies that identify the determinants of successful, sustainable assessment and feedback strategies so that we can better understand factors that promote trainee feedback as a routine feature of training programmes rather than a unique feature of selected programmes only. Long term use may require further modification and simplification of existing methods so as to make them more user-friendly in busy clinical settings where patient care is the first priority and trainee assessment of less importance.

Based on current literature it is apparent that poor faculty participation in formative assessment and feedback strategies is probably the most significant limiting factor currently identified. Why faculty do not routinely engage in trainee assessment and feedback needs to be better understood if we wish to improve the situation. One strategy that may be of benefit would be a reward structure for busy clinicians that appropriately recognises their educational contributions and/or provides them protected time to engage in teaching activities. Another strategy would be to identify a core group of faculty whose only educational job is assessment and formative feedback. Other strategies clearly need to be identified. In any event, these realities need to be addressed before formative assessment is likely to be a routine feature of workplace-based training programmes.

Second, we need to improve the quality of the assessments and feedback given to trainees through a concerted faculty development effort. Current work indicates that feedback rarely results in the formulation of an action plan, a critical component of effective feedback, and only sometimes involves self-assessment by the trainee. Both these issues need to be addressed if feedback is to be "owned" by the trainee and remedial action undertaken to improve performance. In addition, the accuracy and stringency of feedback need to be improved. Innovative strategies to address this important aspect of formative assessment need to be developed.

Finally, the impact of feedback on trainee learning behaviour and performance needs to be determined. To date there is very little information about the strategic use of formative assessment in the workplace context to drive the learning of medical trainees. The need for such data is apparent. Not only do we need to determine the impact of feedback on learning behaviour, but we also need to know what performance-in-the-workplace benefits can be expected to be achieved by successful formative assessment strategies.

"We need studies that identify the determinants of successful, sustainable assessment and feedback strategies so that we can better understand factors that promote trainee feedback as a routine feature of training programmes rather than a unique feature of selected programmes only."

Summary

In the context of the workplace-based education of doctors, there has been concern that trainees are seldom observed, assessed, and given feedback. This has led to increasing interest in a variety of formative assessment methods that require observation and offer the opportunity for feedback, including the mini-clinical evaluation exercise, clinical encounter cards, clinical work sampling, blinded patient encounters, direct observation of procedural skills, case-based discussion, and multisource feedback. The research literature on formative assessment and feedback suggests that it is a powerful means for changing the behaviour of students and trainees.

To enhance the efficacy of the methods of workplace-based assessment, it is critical that the feedback which is provided be consistent with the needs of the learner, focus on important aspects of the performance (while avoiding personal issues), and have a series of characteristics which make it maximally effective. Since faculty play a key role in the successful implementation of formative assessment, strategies to provide training and encourage their participation are critical.

References

Arnold, L., Willoughby, L., Calkins, V. & Eberhart, G. (1981) Use of peer evaluation in the assessment of medical students. Medical Education, 56, pp. 35-41.

Archer, J.C., Norcini, J.J., and Davies, H.A. (2005) Peer review of paediatricians in training using SPRAT. British Medical Journal, 330, pp. 1251-3.

Association of American Medical Colleges. (2004) Medical school graduation questionnaire: all schools report. Available from URL:http://www.aamc.org/data/gq/allschoolsreport/2004.pdf Accessed on 11 April 2007.

Balzer, W.K., Doherty, M.E. & O'Connor, R., Jr. (1989) Effects of cognitive feedback on performance. Psychological Bulletin, 106, pp. 410-433.

Bargh, J.A., Gollwitzer, P.M., Lee-Chai, A., Barndollar, K. & Trotshcel, R. (2001) The automated will: Nonconscious activation and pursuit of behavioural goals. Journal of Personality and Social Psychology, 81, pp. 1014-1027.

Beck, R.S., Daughtridge, R., & Sloane, P.D. (2002). Physician-patient communication in the primary care office: a systematic review. Journal of the American Board of Family Practice, 15, pp 25-38.

Boulet, J.R., McKinley, D.W., Norcini, J.J. & Whelan, G.P. (2002) Assessing the comparability of standardized patient and physician evaluations of clinical skills. Advances in Health Sciences Education, 7, pp. 85-97.

Branch, W.T. & Paranjape, A. (2002) Feedback and reflection: teaching methods for clinical settings. Academic Medicine, 77, pp. 1185-1188.

Burch, V.C., Seggie, J.L., Gary, N.E. Formative assessment promotes learning in undergraduate clinical clerkships. South African Medical Journal 2006; 96: 430-433.

Butler, D.L. & Winnie, P.H. (1995) Feedback and elf-regulated learning: a theoretical synthesis. Review of Educational Research, 65, pp. 245-274.

Craven, R.G., Marsh, H.W. & Debus, R.L. (1991) Effects of internally focused feedback and attributional feedback on enhancement of academic self-concept. Journal of Educational Psychology, 83, pp. 17-27.

Crooks, T.J. (1988) The impact of classroom evaluation practices on students. Review of Educational Research, 58, pp. 438-481.

Day, S.C., Grosso, L.G., Norcini, J.J., Blank, L.L., Swanson, D.B. & Horne, M.H. (1990) Residents' perceptions of evaluation procedures used by their training program. Journal of General Internal Medicine, 5, pp. 421-426.

Daelmans HE, Hoogenboom RJ, Donker AJ, Scherpbier AJ, Stehouwer CD, van der Vleuten CP. (2004) Effectiveness of clinical rotations as a learning environment for achieving competences. Medical Teacher, 26(4) pp. 305-12.

Driessen, E. & van der Vleuten, C. (2000) Matching student assessment to problem-based learning: lessons from experience in a law faculty. Studies in Continuing Education, 22, pp. 235-248.

Durning, S.J., Cation, L.J., Markert, R.J., & Pangaro, L.N. (2002). Assessing the reliability and validity of the mini-clinical evaluation exercise for internal medicine residency training. Academic Medicine, 77, pp. 900-4.

Ende, J. (1983) Feedback in medical education. Journal of American Medical Association, 250, pp. 777-781.

Finlay, K., Norman, G.R., Stolberg, H., Weaver, B., Keane, D.R. (2006). In-training evaluation using hand-held computerized clinical work sampling strategies in radiology residency. Journal of the Canadian Association of Radiology. 57, pp. 232-7.

Frederiksen, N. (1984) The real test bias. Influences on testing and teaching and learning. American Psychologist, 39, pp. 193-202.

Gibbs, G. (1999) Using assessment strategically to change the way students learn. In: S. Brown (Ed.) Assessment Matters in Higher Education. Choosing and using Diverse Approaches. (Buckingham: Society for Research into Higher Education and Open University Press).

Gibbs, G. & Simpson, C. (2004-2005) Conditions under which assessment supports student learning. Learning and Teaching in Higher Education, 1, pp. 3-31.

Gipps, C. (1999) Socio-cultural aspect of assessment. Review of Educational Research, 24, pp. 355-392.

Grondlund, N.E. (1998) Assessment of Student Achievement. 6th ed. (Needham Heights, MA: Allyn and Bacon).

Hampton, J.R., Harrison, M.J.G., Mitchell, J.R.A., Prichard J.S. & Seymour C. (1975) Relative contributions of history-taking, physical examination, and laboratory investigation to diagnosis and management of medical outpatients. British Medical Journal, 2, pp. 486-489.

Harackiewicz, J.M. (1979) The effect of reward contingency and performance feedback on intrinsic motivation. Journal of Personality and Social Psychology, 37, pp. 1352-1363.

Hatala, R., Ainslie, M., Kassen, B.O., Mackie, I., Roberts, J.M. (2006). Assessing the mini-Clinical Evaluation Exercise in comparison to a national specialty examination. Medical Education, 40, pp. 950-6.

Hatala, R. & Norman, G.R. (1999) In-training evaluation during an Internal Medicine clerkship. Academic Medicine, 74, pp. S118-S120.

Hattie, J.A. (1999) Influences on student learning. Inaugural professorial address, University of Auckland, New Zealand. Available from URL:http://www.arts.auckland.ac.nz/staff/index.cfm?P=8650 Accessed on 4 April 2007.

Hattie, J. & Timperley, H. (2007) The power of feedback. Review of Educational Research, 77, pp. 81-112.

Hauer, K.E. (2000) Enhancing feedback to students using the mini-CEX (clinical evalaution exercie). Academic Medicine, 75, p.524.

Herbers, J.E., Noel, G.L., Cooper G.S., Harvey, J., Pangaro, L.N. & Weaver, M.J. (1992) How accurate are faculty evaluations of clinical competence? Journal of General Internal Medicine, 4, pp. 202-208.

Hewson, M.G. & Little, M.L. (1998) Giving feedback in medical education. Verification of recommened techniques. Journal of General Internal Medicine, 13, 111-116.

Holmboe, E.S., Yepes, M., Williams, F. & Huot, S.J. (2004) Feedback and the mini-clinical evaluation exercise. Journal of General Internal Medicine, 19, 558-561.

Holmboe E.S., Hawkins, R.E., Huot, S.J. (2004). Direct observation of competence training: a randomized controlled trial. Annals of Internal Medicine, 140, pp. 874-81.

Isaacson, J.H., Posk, L.K., Litaker, D.G. & Halperin A.K. (1995) Residents' perceptions of the evaluation process. Journal of General Internal Medicine, 10 (suppl.), p.89.

Kalet, A., Earp, J.A. & Kowlowitz, V. (1992) How well do faculty evaluate the interviewing skills of medical students? Journal of General Internal Medicine, 7, pp. 499-505.

Kassebaum, D. G. & Eaglen, R.H. (1999) Shortcoming in the evaluation of studnets' clinical skills and behaviours in medical school. Academic Medicine, 74, pp. 841-849.

Kirch, W. & Schafii, C. (1996) Misdiagnosis at a university hospital in 4 medical eras. Medicine (Baltimore), 75, pp. 29-40.

Kogan, J.R., Bellin, L.M. & Shea, J.A. (2002) Implementation of the mini-CEX to evaluate medical students' clinical skills. Academic Medicine, 77, pp. 1156-1157.

Kogan, J.R. & Hauer, K.E. (2006) Brief report: use of the mini-clinical evalaution exercise in Internal Medicine core clerkships. Journal of General Internal Medicine, 21, pp. 501-502.

Kluger, A.N. & DeNisi, A. (1996) The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. Psychological Bulletin, 119, 254-284.

Little, P., Everitt, H., Williamson, I., Warner, G., Moore, M., Gould, C., Ferrier,K., & Payne, S. (2001). Observational study of effect of patient centredness and positive approach on outcomes of general practice consultations. British Medical Journal, 323, pp. 908-11.

Locke, E.A. & Latham, G.P. (1990) A theory of goal setting and task performance. (Englewood Cliffs, NJ: Prentice Hall).

Luthans, F. & Peterson, S.J. (2004) 360-degree feedback with systematic coaching: empirical analysis suggests a winning combination. Human Resources Management, 42, pp. 243-256.

Maatsch, J.L., Huang, R., Downing, S., & Barker, B. (1983) Predictive validity of medical specialist examinations. Final report for Grant HS 02038-04, National Center of Health Services Research. Office of Medical Education Research and Development, Michigan State University, East Lansing, MI.

Marton, F., Dall'Alba, G. & Beaty, E. (1993) Conceptions of learning. International Journal of Educational Research, 19, pp. 277-300.

McLeod, P.J. & Meagher, T.W. (2001) Educational benefits of blinding students to information acquired and management plans generated by other physicians. Medical Teacher, 23, pp. 83-85.

National Health Service. Modernising Medical Careers: Foundation Programmes. Available from URL:http//:www.mmc.nhs.uk/pages/foundation Accessed on 7 April 2007.

Newble, D.I. & Jaeger, K. (1983) The effect of assessments and examinations on the learning of medical students. Medical Education, 17, pp. 165-171.

Noel, G.L., Herbers, J.E., Caplow, M.P. Cooper, G.S., Pangaro L.N. & Harvey, J. (1992) How well do Internal Medicine faculty members evaluate the clinical skills of residents? Journal of General Internal Medicine, 117, pp. 757-765.

Norcini, J.J., Blank, L.L., Arnold, G.K., and Kimball, H.R. (1995). The mini-CEX (clinical evaluation exercise): A preliminary investigation. Annals of Internal Medicine, 123, pp. 795-799.

Norcini, J.J., Blank, L.L., Duffy, F.D. and Fortna, G. (2003). The mini-CEX: A method for assessing clinical skills. Annals of Internal Medicine, 138, pp. 476-481.

Norcini, J.J. (2007) Workplace-based assessment in clinical training. In Swanwick T. (Ed.): Understanding Medical Education series. Association for the Study of Medical Education Edinburgh, UK.

Norman, G.R., Davis, D., Painvin, A., Lindsay, E., Rath, D., Ragbeer, M. (1989). Comprehensive assessment of clinical competence of family/general physicians using multiple measures. Proceedings of the Research in Medical education Conference, pp 75-79.

Paukert, J.L., Richards, M.L. & Olney, C. (2002) An encounter card system for increasing feedback to students. The American Journal of Surgery, 183, pp. 300-304.

Peterson, M.C., Holbrook, J.H., Hales, D.V., Smith, N.L. & Staker, L.V. (1992) Contributions of the history, physical examination and laboratory investigation in making medical diagnoses. The Western Journal of Medicine, 156, pp. 163-165.

Ramsey, P., Carline, J.D., Inui, T.S., Larson, E.B., LoGerfo, J.P. and Wenrich, M.D.. (1989). Predictive validity of certification by the American Board of Internal Medicine. Annals of Internal Medicine, 110, pp. 719-726.

Ramsey, P., Wenrich, M.D., Carline, J.D., Inui, T.S., Larson, E.B., and LoGerfo, J.P. (1993). Use of peer ratings to evaluate physician performance. Journal of the American Medical Association, 269, pp. 1655-1660.

Richards ML, Paukert JL, Downing SM, Bordage G. (2007). Reliability and usefulness of clinical encounter cards for a third-year surgical clerkship. Journal of Surgical Research, 140, pp. 139-48.

Sadler, R. (1989) Formative assessment and the design of instructional systems. Instructional Science, 18, pp. 119-144.

Shepard, L.A. (2000) The role of assessment in a learning culture. Educational Researcher, 29, pp. 4-14.

Small, P.A., Stevens, B. & Duerson, M.C. (1993) Issues in medcial education: basic problems and potential solutions. Academic Medicine, 68, pp. S89-S98.

Solomon, D.J., Reinhart, M.A., Bridgham, R.G., Munger, B.S., Starnaman, S. (1990). An assessment of an oral examination format for evaluating clinical competence in emergency medicine. Academic Medicine, 65 (Suppl), pp. S43-4.

Stillman, P.L., Haley, H-L., Regan, M.B. & Philbin. (1991) M.M. Positive effects of a clinical performance assessment programme. Academic Medicine, 66, pp. 481-483.

Swanson, D.B., Norman, G.R. & Linn, R.L. (1995) Performance-based assessment: lessons from the health professions. Educational Researcher, 24, pp. 5-11.

Thompson, T. (1998) Metamemory accuracy: effects of feedback and the stability of individual differences. American Journal of Psychology, 111, pp. 33-42.

Turnbull, J., MacFayden, J., van Barneveld & C., Norman G. (2000) Clinical works sampling. A new approach to the problem of in-training evaluation. Journal of General Internal Medicine, 15, 556-561.

Van der Vleuten, C.P.M. (1996) The assessment of professional competence: developments, research and practical implications. Advances in Health Sciences Education, 1, pp. 41-61.

Veloski, J., Boex, J.R., Grasberger, J., Evans, A.,& Wolfson, D.B. (2006). Systematic review of the literature on assessment, feedback, and physicians' clinical performance: BEME Guide No 7. Medical teacher, 28, pp117-128.

Walker, J. (2001). The managerial mentor-leading productive learning in the workplace: an integral view. University of Technology Sydney Research Centre Vocational Education & Training. Productive learning seminar series, November, p.3. Available from URL:http://www.oval.uts.edu.au/working_papers/2002WP/0209walker.pdf Accessed on 4 April 2007.

Wenrich, M.D., Carline J.D., Giles, L.M. & Ramsey, P.G. (1993) Ratings of the performance of practicing internists by hospital-based registered nurses. Academic Medicine, 68, pp.680-687.

Whitehouse, A., Waltzman, M., & Wall, D. (2002) Pilot study of 3600 assessment of personal skills to inform record of in-training assessments for senior house officers. Hospital Medicine, 63, pp. 172-175.

Winnie, P.H. & Butler, D.L. (1994) Student cognition in learning from teaching. In: Husen & T. Postlewaite (Eds.), International Encyclopedia of Education, pp. 5738-5745 (Oxford, UK: Pergamon).

Wragg, A., Wade, W., Fuller, G., Cowan, G. & Mills, P. (2003) Assessing the performance of specialist registrars. Clinical Medicine, 3, pp. 131-134.

Notes

AMEE Guides – the new series

Welcome to the new series of AMEE Guides.

- The AMEE guides cover important topics in medical and healthcare professions education and provide information, practical advice and support. We hope that they will also stimulate your thinking and reflection on the topic.

- The guides have been logically structured for ease of reading and contain useful take-home messages. Text boxes highlight key points and examples in practice.

- Each page in the guide provides a column for your own personal annotations, stimulated either by the text itself or the additional quotation.

- Sources of further information on the topic are provided in the reference list and bibliography.

- The guides are designed for use by individual teachers to inform their practice and can be used to support staff development programmes.

- For each guide, supplements will be prepared that provide additional examples and contributions relating to the topic.

'Living Guides'

An important feature of this new guide series is the concept of supplements, which will provide for you a continuing source of information on the topic. Published supplements will be emailed to those who have purchased the guide and they will be included in the text of future issues. To register to receive the supplements, please email the AMEE Office: amee@dundee.ac.uk. We suggest you print out the supplements and add these to the back of the guide, the outer cover of which has been designed to accommodate A4 inserts.

You may also wish to contribute through the supplements to the further development of this AMEE guide. If you have experience in the area, with examples or guidelines from practice that might be of value to others, we would be pleased to hear from you. Submissions for consideration for inclusion as a guide supplement should fit onto a double-sided A4 sheet, approximately 700-1000 words. They should be sent to Pat Lilley, AMEE Administrator, p.m.lilley@dundee.ac.uk Selected submissions will be published in Medical Teacher. If you have any queries relating to possible contributions please contact the Guides Series Editor, Professor Trevor Gibbs (tjg.gibbs@gmail.com).

Other guides in the new series

A full list of topics in this exciting new guide series is available on the AMEE website. We hope you enjoy reading them and find the supplements a source of information on the topic.
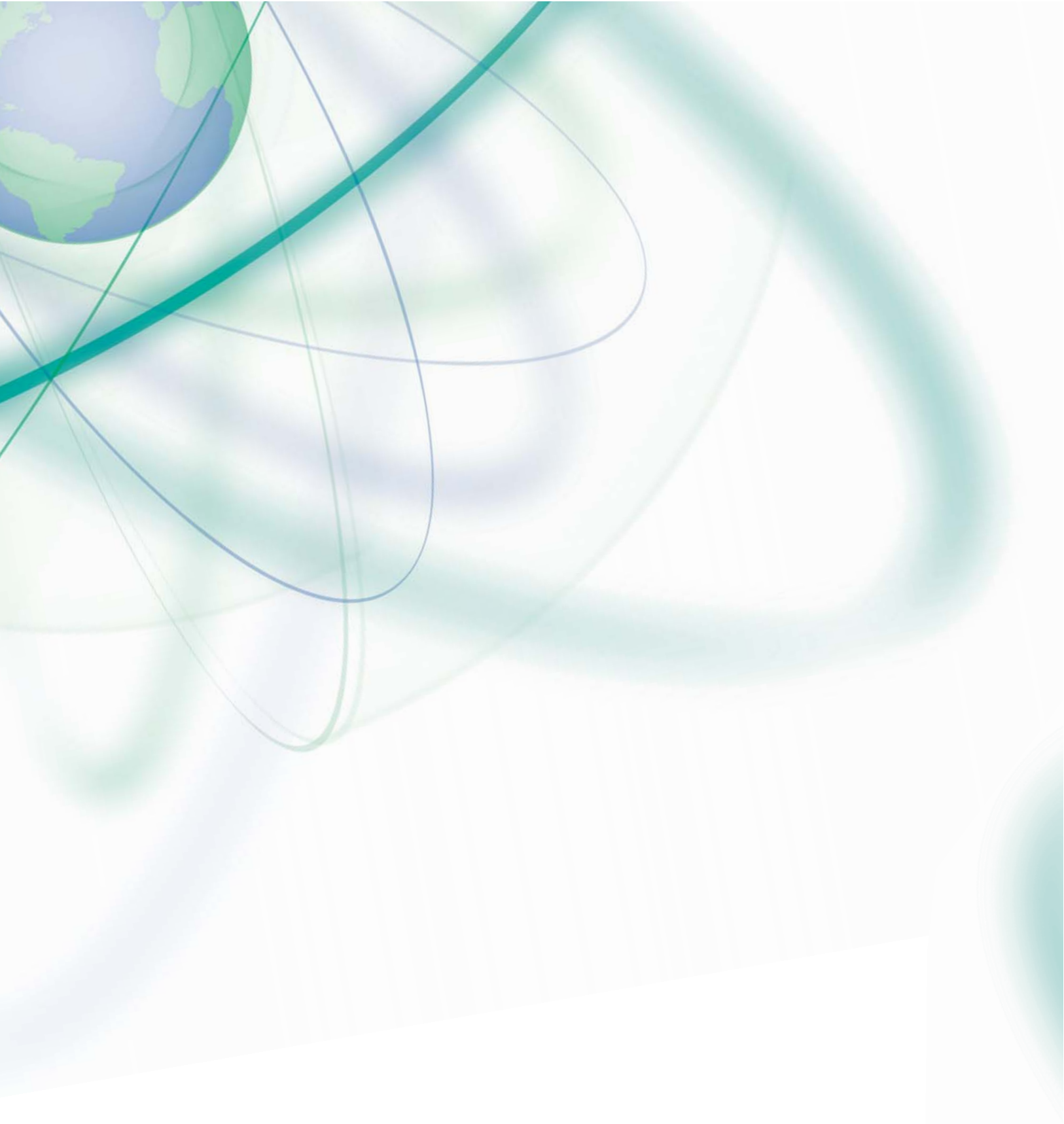
For information about other guides in the series and how to
order copies, please see the AMEE website: www.amee.org