

Impact of workplace based assessment on doctors' education and performance: a systematic review

Alice Miller, National Institute for Health Research academic clinical fellow,¹ Julian Archer, National Institute for Health Research academic clinical lecturer¹

Peninsula College of Medicine and Dentistry, University of Plymouth, Plymouth PL4 8AA, UK

Correspondence to: A Miller
alice.miller@pms.ac.uk

Cite this as: *BMJ* 2010;341:c5064
doi:10.1136/bmj.c5064

ABSTRACT

Objective To investigate the literature for evidence that workplace based assessment affects doctors' education and performance.

Design Systematic review.

Data sources The primary data sources were the databases Journals@Ovid, Medline, Embase, CINAHL, PsycINFO, and ERIC. Evidence based reviews (Bandolier, *Cochrane Library*, DARE, HTA Database, and NHS EED) were accessed and searched via the Health Information Resources website. Reference lists of relevant studies and bibliographies of review articles were also searched.

Review methods Studies of any design that attempted to evaluate either the educational impact of workplace based assessment, or the effect of workplace based assessment on doctors' performance, were included. Studies were excluded if the sampled population was non-medical or the study was performed with medical students. Review articles, commentaries, and letters were also excluded. The final exclusion criterion was the use of simulated patients or models rather than real life clinical encounters.

Results Sixteen studies were included. Fifteen of these were non-comparative descriptive or observational studies; the other was a randomised controlled trial. Study quality was mixed. Eight studies examined multisource feedback with mixed results; most doctors felt that multisource feedback had educational value, although the evidence for practice change was conflicting. Some junior doctors and surgeons displayed little willingness to change in response to multisource feedback, whereas family physicians might be more prepared to initiate change. Performance changes were more likely to occur when feedback was credible and accurate or when coaching was provided to help subjects identify their strengths and weaknesses. Four studies examined the mini-clinical evaluation exercise, one looked at direct observation of procedural skills, and three were concerned with multiple assessment methods: all these studies reported positive results for the educational impact of workplace based assessment tools. However, there was no objective evidence of improved performance with these tools.

Conclusions Considering the emphasis placed on workplace based assessment as a method of formative

performance assessment, there are few published articles exploring its impact on doctors' education and performance. This review shows that multisource feedback can lead to performance improvement, although individual factors, the context of the feedback, and the presence of facilitation have a profound effect on the response. There is no evidence that alternative workplace based assessment tools (mini-clinical evaluation exercise, direct observation of procedural skills, and case based discussion) lead to improvement in performance, although subjective reports on their educational impact are positive.

INTRODUCTION

The assessment of clinical performance in medicine is important but challenging. Historically, assessments have been implicit, unstandardised, and based on holistic or subjective judgments (the apprenticeship model).¹ However, recent reforms in postgraduate medical education^{2,3} have brought new systems for the assessment of competence and performance.

Workplace based assessment is one of these systems. Workplace based assessment refers to "the assessment of day-to-day practices undertaken in the working environment"⁴—or, more simply, workplace based assessment is an "assessment of what doctors actually do in practice."⁵ Although many forms of assessment can be used to show a doctor's knowledge or competence, there is evidence that competence does not reliably predict performance in clinical practice⁶; one major advantage of workplace based assessment is its ability to evaluate performance in context.⁷

Another strength of workplace based assessment is its formative potential. A recently published guideline for the implementation of workplace based assessment emphasises the importance of using such tools as assessments for learning rather than solely as assessments of learning.⁸ The critical element required to achieve this is the provision of feedback from assessor to trainee, enabling the trainee to steer his or her learning towards desired outcomes.⁹ There is now convincing evidence that systematic feedback delivered by a credible source can change clinical performance,¹⁰ although there are many complexities that influence the effectiveness of feedback in practice.¹¹

Many different workplace based assessment methods exist, all designed to assess different aspects of performance. Commonly, assessment tools will fit into one of the following categories⁵:

- Observation of clinical activities, such as the mini-clinical evaluation exercise and direct observation of procedural skills
- Discussion of clinical cases, such as chart stimulated recall and the case based discussion
- Feedback from peers, coworkers, and patients, collected by survey and usually called multisource feedback; sometimes known as 360° assessment. Tools for gathering this data include the mini-assessment tool (mini-PAT), team assessment of behaviours (TAB), and the patient satisfaction questionnaire (PSQ).

These tools have been described in more detail elsewhere.^{9,12-14} Certain aspects of their utility¹ (particularly their reliability, validity, and acceptability) have been scrutinised over the past few years,^{9,14} but there is still relatively little known about their educational impact.

It is tempting to suggest that, because workplace based assessment requires the provision of feedback, and feedback can lead to learning and improved performance, the implementation of such assessment strategies will have a positive impact on doctors' learning and performance. However, despite the considerable weight placed on them in postgraduate training, there is little information in the medical education literature to support this claim.

The aim of this study was therefore to perform a systematic review of the literature to investigate the educational impact of workplace based assessment in an attempt to answer the question: "What is the evidence that workplace based assessment affects physician education and performance?"

METHODS

Search strategy

The primary data sources for this review were the electronic databases Journals@Ovid (English language only, 1996–February 2010), Medline (1950–February 2010), Embase (1980–February 2010), CINAHL (1981–February 2010), PsycINFO (1806–February 2010), and ERIC (1966–February 2010). Evidence based reviews (Bandolier, *Cochrane Library*, DARE, H TA Database, and NHS EED) were accessed and searched via the Health Information Resources website, www.library.nhs.uk/default.aspx (formerly the National Library for Health).

The search terms (in English only) were

- "Workplace based assessment" or "mini clinical evaluation exercise" or "direct observation of procedural skills" or "case based discussion" or "multisource feedback" (all fields)
- Performance or "educational impact" (all fields)
- Evaluation (all fields)
- Medicine (all fields)

Results from the four searches were combined with "AND" then duplicated results were removed. The remaining citations were displayed and examined. We decided to limit the search to the terms mini-clinical evaluation exercise, direct observation of procedural skills, case based discussion, and multisource feedback because these four tools are in common use internationally. Otherwise, terms used were kept as broad as possible to maximise the chance of finding relevant articles. Hyphens and abbreviations were not used in case they limited the search.

In addition, we searched reference lists of relevant studies and bibliographies of review articles.

Inclusion and exclusion criteria

Eligibility judgments were made by a single author (AM), with consensus from the second author (JA), on the basis of information found in the article's title, abstract, or full text if necessary. Studies were included in the review if they met the following criteria:

- Any study attempting to evaluate the educational impact of workplace based assessment (mini-clinical evaluation exercise, direct observation of procedural skills, case based discussion, or multisource feedback)
- Any study attempting to evaluate the effect of workplace based assessment (mini-clinical evaluation exercise, direct observation of procedural skills, case based discussion, or multisource feedback) on doctors' performance.

Studies were excluded if the sampled population was non-medical or the study was performed with medical students. Review articles, commentaries, and letters were also excluded. The final exclusion criterion was the use of simulated patients or models rather than real life clinical encounters.

Data extraction

Data from eligible articles were extracted into a form to compare the studies. This procedure was performed by both authors independently, with disagreements being resolved by discussion. Column headings were

- Country of origin
- Study population and sample size
- Study design
- Tool(s) being investigated
- Study quality
- Outcome.

We assessed study quality using a series of quality indicators developed by Buckley et al¹⁵ as a Best Evidence Medical Education (BEME) guide (see box 1). We considered studies to be of higher quality if they met seven or more of these 11 indicators.

As a means of evaluating outcome, we applied Barr's adaptation of Kirkpatrick's four level evaluation model (see box 2) to the results of each study.¹⁶ Levels of evaluation were included in the outcome column of the data extraction table.

Box 1: BEME quality indicators developed by Buckley et al¹⁵

- Research question*—Is the research question(s) or hypothesis clearly stated?
- Study subjects*—Is the subject group appropriate for the study being carried out?
- Data collection methods*—Are the methods used reliable and valid for the research question and context?
- Completeness of data*—Have subjects dropped out? Is the attrition less than 50%? Is the questionnaire response rate acceptable?
- Control for confounding*—Have multiple factors or variables been removed or accounted for where possible?
- Analysis of results*—Are the statistical or other methods of results analysis used appropriate?
- Conclusions*—Is it clear that the data justify the conclusions drawn?
- Reproducibility*—Could the study be repeated by other researchers?
- Prospective*—Does the study look forwards in time rather than backwards?
- Ethical issues*—Were all relevant ethical issues addressed?
- Triangulation*—Were results supported by data from more than one source?

RESULTS**Search results**

Our initial search was carried out using the Ovid database because of its good coverage of medical education literature; this yielded 201 articles. We screened titles and abstracts, leading to the exclusion of 163 articles. The remaining 38 articles were read in full, but 27 of these did not fit the inclusion criteria, leaving 11 studies for inclusion in the review.

We then performed the same search via the Health Information Resources website using Medline (15 studies identified), Embase (11 studies identified), CINAHL (60 studies identified), and PsycINFO (39 studies identified). No additional studies were identified from the evidence based reviews section. Removal of duplicates produced 114 articles, the titles and abstracts of which we screened and cross referenced with the initial Ovid search results: 110 either did not fit the inclusion criteria or had already been identified, leaving four new articles for inclusion in the review.

The ERIC database highlighted 14 articles, but all had been previously identified. Manual searching of reference lists identified one additional article of relevance, taking the total number of included articles to 16.¹⁷⁻³² The figure summarises the results of the search strategy.

General findings

Details of the included studies and a summary of the data extracted are listed in the web extra table on bmj.com. Most of the studies were conducted in the United Kingdom and Canada, with smaller numbers originating from New Zealand, Australia and the United States. Study populations consisted of doctors from all levels of training and from different specialties, including primary and secondary care. Fifteen of the 16 included studies were non-comparative descriptive or observational studies, and one was a randomised controlled trial. Study quality was mixed. Of the included articles, 11 were graded as higher quality

and five as lower quality according to BEME quality indicators (see web extra table). Eight studies examined multisource feedback,¹⁷⁻²⁴ four concentrated on the mini-clinical evaluation exercise,²⁵⁻²⁸ one investigated direct observation of procedural skills,²⁹ and three looked at multiple assessment methods.³⁰⁻³²

Multisource feedback

As part of a larger study investigating the reliability and feasibility of various different workplace based assessment methods in general practice training, Murphy et al²⁰ asked study participants to rate the educational impact of multisource feedback on a 7-point Likert scale. The mean score was 4.2, indicating that most doctors felt multisource feedback held educational value (Kirkpatrick level 1). Other studies went further, attempting to show that multisource feedback could lead to modifications in attitudes (Kirkpatrick level 2a)^{18 19} or even changes in behaviour (Kirkpatrick level 3).^{21 22 24}

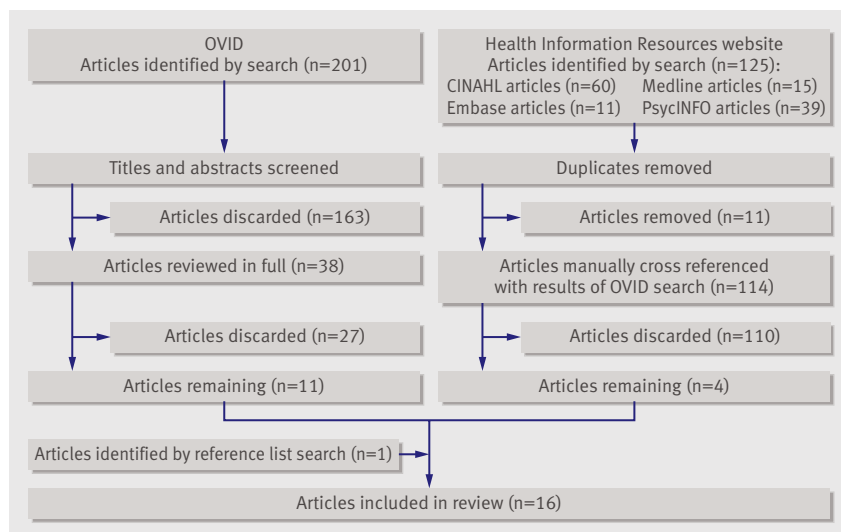
However, in terms of tool effectiveness, results are mixed. A questionnaire of 249 foundation year 1 doctors¹⁸ revealed that nearly a third of trainees did not anticipate changing in response to multisource feedback. Similarly, a group of fully qualified surgeons were unlikely to make practice changes in response to feedback data, even if their multisource feedback scores revealed a need to consider change.¹⁹ On a more positive note, a survey of 113 family physicians²¹ showed that 61% had either made practice changes or were planning to make them in response to their multisource feedback data. More detailed, focus group data from a small sample of these doctors²² revealed that feedback is useful only if it is perceived to be accurate and credible; feedback perceived as negative and inaccurate is much less likely to lead to practice improvement.

One prospective longitudinal study²⁴ collected multisource feedback data from 250 family physicians on two separate occasions five years apart and found small to moderate improvements in scores the second time. However, the authors were unable to conclude that this performance improvement was due to the multisource feedback.

Brinkman et al¹⁷ carried out a randomised controlled trial to determine whether multisource feedback could lead to improvements in communication skills and professional behaviours in paediatric residents. As well as receiving a feedback report from collated multisource feedback data, participants in the intervention

Box 2: Barr's adaptation of Kirkpatrick's evaluation model¹⁶

- Level 1*—Learner's reactions
- Level 2a*—Modification of attitudes and perceptions
- Level 2b*—Acquisition of knowledge and skills
- Level 3*—Change in behaviour
- Level 4a*—Change in organisational practice
- Level 4b*—Benefits to clients or patients



Results of the search strategy for articles

group were also obliged to fill in a self assessment form and to take part in a tailored coaching session to help them identify their strengths and weaknesses. Participants in the control group received standard feedback only. After five months, the multisource feedback group showed significant improvements in “communicating effectively with the patient and family (35%; 95% confidence interval, 11.0%-58.0%), timeliness of completing tasks (30%; 95% confidence interval, 7.9%-53.0%), and demonstrating responsibility and accountability (26%; 95% confidence interval, 2.9%-49.0%),” but only when rated by nursing staff.¹⁷ However, it is not clear whether the same performance improvements would have occurred without the tailored coaching sessions provided.

Mini-clinical evaluation exercise

The four studies evaluating the mini-clinical evaluation exercise were all concerned with its educational impact as a formative assessment tool. A qualitative study from Canada²⁵ investigated internal residents’ perceptions of the exercise as an educational tool; most agreed that it had had a positive educational impact (Kirkpatrick level 1). The focus group participants also highlighted the point that use of the mini-clinical evaluation exercise as an assessment limited its value as an educational tool.

Nair et al studied the reliability, feasibility, and acceptability of the mini-clinical evaluation exercise in a group of international medical graduates,²⁶ and found that nearly half were either satisfied or very satisfied with the exercise as a tool for learning (Kirkpatrick level 1).

Two studies from New Zealand looked at the educational impact of the mini-clinical evaluation exercise in anaesthesia training.^{27,28} Survey data revealed that the large majority of trainees (and their assessors) felt that the evaluation exercise improved the frequency and quality of feedback offered (Kirkpatrick level 1).²⁷ Focus group and interview data built on these findings,

suggesting that the mini-clinical evaluation exercise promoted educational interaction and improved training quality (Kirkpatrick level 1).²⁸

We found no studies looking at the effect of the mini-clinical evaluation exercise on doctors’ performance.

Direct observation of procedural skills

An observational survey describing the implementation of direct observation of procedural skills, mini-clinical evaluation exercise, and multisource feedback in a London hospital²⁹ provides some data on the educational impact of direct observation of procedural skills. A feedback survey returned by 25 of the 27 preregistration house officers completing the assessments revealed that most (70%) felt that direct observation helped to improve clinical skills (Kirkpatrick level 2b). Furthermore, 65% agreed with the statement, “I think that undertaking direct observation of procedural skills will improve my future career.” However, this study was graded as lower quality according to BEME indicators, and there was no evidence in this study (or any others) that direct observation of procedural skills leads to objective performance improvement.

Multiple assessment methods

Three studies looked at the impact of multiple assessment methods on education and training. A large survey collected the opinions of 539 surgical trainees on the Intercollegiate Surgical Curriculum Programme,³⁰ an online portfolio that is used to administer various workplace based assessments (including mini-clinical evaluation exercise, case based discussion, direct observation of procedural skills, and multisource feedback). Over 60% of survey respondents felt that the programme sometimes or frequently impacted adversely on training opportunities, as a result of the time needed to complete the assessments. More than 90% stated that the programme had a neutral or negative impact on their training overall (Kirkpatrick level 1).

A questionnaire of 95 foundation year 2 doctors exploring their experiences of the foundation programme portfolio was slightly more positive.³¹ Most felt that the portfolio was effective in helping them to achieve their educational requirements (Kirkpatrick level 1), but some felt that its success as an educational tool was limited by lack of understanding of its contents and purpose (particularly by educational supervisors).

An observational study to evaluate the reliability and feasibility of workplace based assessment for assessing medical registrars also gave positive results in terms of educational impact.³² Participants completed a questionnaire about their experiences of workplace based assessment, and the large majority felt that mini-clinical evaluation exercise, direct observation of procedural skills, and multisource feedback were helpful in aiding personal development (Kirkpatrick level 1). There were also positive free-text comments about the ability of the assessments to provide a basis for feedback, although many also found them to be time consuming and a considerable administrative workload.

Again, there were no studies investigating the impact of multiple assessment methods on performance.

DISCUSSION

This systematic review brings together the available evidence concerning the educational impact of workplace based assessment and its ability to change doctors' performance. Considering the emphasis now placed on workplace based assessment as a method of formative performance assessment, there are surprisingly few published articles exploring these areas, and the strength of the findings is questionable.

The strongest evidence for workplace based assessment improving performance comes from studies examining multisource feedback. Work reported in the psychology literature has shown that multisource feedback can lead to small improvements in performance over time,³³ and a 10 year old study of medical education also showed that doctors exposed to specific feedback from peers, coworkers, and patients can use the data to inform changes in their practice.³⁴ The studies we reviewed show conflicting evidence, however. Although some junior doctors¹⁸ and most surgeons¹⁹ displayed little willingness to change in response to multisource feedback, family physicians seemed more prepared to initiate performance changes.²¹ This variability may be due to individual differences; it is already known that performance improvement is more likely to occur when feedback indicates a need for change, when recipients have a positive view of feedback, and when they believe that change is feasible.³³

The single randomised controlled trial in our review¹⁷ attempted to show improved performance in the intervention group allocated to multisource feedback, but the positive results seen might have been due to the coaching session that was also part of the intervention. The positive influence of facilitation in the effectiveness of multisource feedback has recently been established,^{11,35} especially in the context of negative feedback.³⁶

It seems, therefore, that multisource feedback can lead to improved performance, but individual factors, the context of the feedback, and the presence (or absence) of facilitation can have a profound effect on the magnitude of the response.

We were unable to unearth any clear evidence to show that the mini-clinical evaluation exercise, direct observation of procedural skills, or case based discussion can lead to improvements in performance. The studies examining the mini-clinical evaluation exercise and multiple assessment methods showed largely positive results in terms of learner satisfaction but could not show changes in attitudes, skills, knowledge, or behaviour. The study of the impact of direct observation of procedural skills³⁰ revealed that some house officers felt it could improve their clinical skills, but this evidence has not been captured objectively, and participant numbers were small. A previous systematic review investigating tools for direct observation and assessment of clinical skills found similarly few studies describing educational outcome.³⁷

Limitations

Most of the articles included in this review were non-comparative descriptive or observational studies. Strength of findings may be limited by the uncontrolled nature of the studies, but given the methodological difficulties of evaluating topics such as educational impact and doctor performance,³⁸ descriptive and observational studies can still provide useful information. Indeed, some of the strongest evidence for improved performance after workplace based assessment comes from detailed focus group data.²² The single randomised controlled trial we identified attempted to establish causality ("multisource feedback causes performance improvement"),¹⁷ but, as discussed above, the results are undoubtedly affected by confounding factors.

Methodological rigour is clearly apparent in some articles, especially those aiming to evaluate multiple facets of workplace based assessment,^{20,32} but, because the focus here tends to be on reliability and feasibility, they may be less suitable for gathering data about educational impact or performance change.

Quality is also affected by the voluntary nature of participation in most of the studies. Potentially biased³⁰ or highly motivated²¹ study populations can lead to profoundly different results. The reliance on self reporting and the small study populations in most of the studies also limit the quality and strength of their findings.

Our review methodology also has its limitations. Although our database search was extensive, we did not review the grey literature and so may have missed some relevant studies. The Ovid database search was also limited to English language publications, so there may have been publication bias. The search may also have been limited by the terms used (for example, multisource feedback is also known as 360° feedback, mini-peer assessment tool, and team assessment of behaviours, but these terms were not included in the search).

Conclusions

This review has highlighted once again the need for further research in the area of formative performance assessment: the increasing use of workplace based assessment methods in postgraduate medical training and recertification should provide fertile ground for this work. Serious consideration needs to be given to the use of study designs that are able to show conclusive links between workplace based assessment and performance improvement. So often workplace based assessment has been implemented wholesale and evaluation has subsequently and understandably focused on feasibility and self reported outcomes. We need to move to an interventionist, experimental model to establish whether workplace based assessment makes a difference. Future studies will need to be ambitious, not only in size to show significant change but over extended periods to expose matched groups of doctors to different interventions. This will require collaboration within and across nations.

WHAT IS ALREADY KNOWN ON THIS SUBJECT

Workplace based assessment involves feedback on performance of doctors in their everyday activities

It purports to provide useful information to both doctors and to their training and regulatory bodies, and is often reported to support educational impact and learning

WHAT THIS STUDY ADDS

This review found little evidence in the literature to date for workplace based assessment as an educational initiative

There is limited evidence that multisource feedback may lead to improved performance, but factors such as the context of the feedback and the presence of facilitation seem to have a profound effect on the response

Further avenues for future work are also clearly signposted from here. A focus on assessment programmes to show how workplace based assessment instruments can be used together would be of great practical benefit. The role of facilitation in workplace based assessment, and the extent of its involvement in performance improvement, must also be fully investigated. Finally, we need to discover whether formative assessment strategies such as workplace based assessment can reach Kirkpatrick's highest levels of change, leading to improvements in care delivery and patient outcome.

Funding: None.

Competing interests: All authors have completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) (URL) and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous 3 years; no other relationships or activities that could appear to have influenced the submitted work.

Data sharing: No additional data available.

- 1 Van der Vleuten CPM. The assessment of professional competence: development, research and practical implications. *Adv Health Sci Educ* 1996;1:41-67.
- 2 Modernising Medical Careers. Medical specialty training (England). 2010. www.mmc.nhs.uk.
- 3 Scheele F, Teunissen P, Van Luijk S, Heineman E, Fluit L, Mulder H, et al. Introducing competency-based postgraduate medical education in the Netherlands. *Med Teach* 2008;30:248-53.
- 4 Postgraduate Medical Education and Training Board Workplace Based Assessment Subcommittee. *Workplace based assessment*. Postgraduate Medical Education and Training Board, 2005.
- 5 Swanwick T, Chana N. Workplace-based assessment. *Br J Hosp Med (Lond)* 2009;70:290-3.
- 6 Rethans JJ, Norcini JJ, Baron-Maldonado M, Blackmore D, Jolly BC, LaDuca T, et al. The relationship between competence and performance: implications for assessing practice performance. *Med Educ* 2002;36:901-9.
- 7 Miller GE. The assessment of clinical skills/competence/performance. *Acad Med* 1990;65:S63-7.
- 8 General Medical Council, Academy of Medical Royal Colleges. *Workplace based assessment: a guide for implementation*. 2010. www.gmc-uk.org/Workplace_Based_Assessment.pdf_31300577.pdf.
- 9 Norcini J, Burch V. Workplace-based assessment as an educational tool: AMEE Guide No 31. *Med Teach* 2007;29:855-71.
- 10 Veloski J, Boex JR, Grasberger MJ, Evans A, Wolfson DB. Systematic review of the literature on assessment, feedback and physicians' clinical performance: BEME Guide No 7. *Med Teach* 2006;28:117-28.
- 11 Archer JC. State of the science in health professional education: effective feedback. *Med Educ* 2010;44:101-8.
- 12 Carr S. The foundation programme assessment tools: an opportunity to enhance feedback to trainees? *Postgrad Med J* 2006;82:576-9.
- 13 Davies H, Archer J, Heard S, Southgate L. Assessment tools for foundation programmes - a practical guide. *BMJ Careers* 2005;330:195-6.
- 14 Norcini JJ. Workplace-based assessment in clinical training. In: Swanwick T, ed. *Understanding medical education*. Association for the Study of Medical Education, 2007.
- 15 Buckley S, Coleman J, Davison I, Khan KS, Zamora J, Malick S, et al. The educational effects of portfolios on undergraduate student learning: a Best Evidence Medical Education (BEME) systematic review. BEME Guide No 11. *Med Teach* 2009;31:282-98.
- 16 Barr H, Freeth D, Hammick M, Koppel I, Reeves S. *Evaluations of interprofessional education: a United Kingdom review of health and social care*. CAIPE/BERA, 2000.
- 17 Brinkman WB, Geraghty SR, Lanphear BP, Khoury JC, Gonzalez del Rey JA, Dewitt TG, et al. Effect of multisource feedback on resident communication skills and professionalism: a randomized controlled trial. *Arch Pediatr Adolesc Med* 2007;161:44-9.
- 18 Burford B, Illing J, Kergon C, Morrow G, Livingston M. User perceptions of multi-source feedback tools for junior doctors. *Med Educ* 2010;44:165-76.
- 19 Lockyer J, Violato C, Fidler H. Likelihood of change: a study assessing surgeon use of multisource feedback data. *Teach Learn Med* 2003;15:168-74.
- 20 Murphy DJ, Bruce DA, Mercer SW, Eva KW. The reliability of workplace-based assessment in postgraduate medical education and training: a national evaluation in general practice in the United Kingdom. *Adv Health Sci Educ Theory Pract* 2009;14:219-32.
- 21 Sargeant JM, Mann KV, Ferrier SN, Langille DB, Muirhead PD, Hayes VM, et al. Responses of rural family physicians and their colleague and coworker raters to a multi-source feedback process: a pilot study. *Acad Med* 2003;78:S42-4.
- 22 Sargeant J, Mann K, Ferrier S. Exploring family physicians' reactions to multisource feedback: perceptions of credibility and usefulness. *Med Educ* 2005;39:497-504.
- 23 Violato C, Lockyer J, Fidler H. Multisource feedback: a method of assessing surgical practice. *BMJ* 2003;326:546-8.
- 24 Violato C, Lockyer JM, Fidler H. Changes in performance: a 5-year longitudinal study of participants in a multi-source feedback programme. *Med Educ* 2008;42:1007-13.
- 25 Malhotra S, Hatala R, Coumeya CA. Internal medicine residents' perceptions of the mini-clinical evaluation exercise. *Med Teach* 2008;30:414-9.
- 26 Nair BR, Alexander HG, McGrath BP, Parvathy MS, Kilsby EC, Wenzel J, et al. The mini clinical evaluation exercise (mini-CEX) for assessing clinical performance of international medical graduates. *Med J Aust* 2008;189:159-61.
- 27 Weller JM, Jolly B, Misur MP, Merry AF, Jones A, Crossley JG, et al. Mini-clinical evaluation exercise in anaesthesia training. *Br J Anaesth* 2009;102:633-41.
- 28 Weller JM, Jones A, Merry AF, Jolly B, Saunders D. Investigation of trainee and specialist reactions to the mini-clinical evaluation exercise in anaesthesia: implications for implementation. *Br J Anaesth* 2009;103:524-30.
- 29 Morris A, Hewitt J, Roberts CM. Practical experience of using directly observed procedures, mini clinical evaluation examinations, and peer observation in pre-registration house officer (FY1) trainees. *Postgrad Med J* 2006;82:285-8.
- 30 Pereira EA, Dean BJ. British surgeons' experiences of mandatory online workplace-based assessment. *J R Soc Med* 2009;102:287-93.
- 31 Ryland I, Brown J, O'Brien M, Graham D, Gillies R, Chapman T, et al. The portfolio: how was it for you? Views of F2 doctors from the Mersey Deanery Foundation Pilot. *Clin Med* 2006;6:378-80.
- 32 Wilkinson JR, Crossley JG, Wragg A, Mills P, Cowan G, Wade W. Implementing workplace-based assessment across the medical specialties in the United Kingdom. *Med Educ* 2008;42:364-73.
- 33 Smither JW, London M, Reilly RR. Does performance improve following multisource feedback? A theoretical model, meta-analysis and review of empirical findings. *Pers Psychol* 2005;58:33-66.
- 34 Fidler H, Lockyer JM, Toews J, Violato C. Changing physicians' practices: the effect of individual feedback. *Acad Med* 1999;74:702-14.
- 35 Overeem K, Wollersheim H, Driessen E, Lombarts K, van de Ven G, Grol R, et al. Doctors' perceptions of why 360-degree feedback does (not) work: a qualitative study. *Med Educ* 2009;43:874-82.
- 36 Sargeant J, Mann K, Sinclair D, Van der Vleuten C, Metsemakers J. Understanding the influence of emotions and reflection upon multisource feedback acceptance and use. *Adv Health Sci Educ Theory Pract* 2008;13:275-88.
- 37 Kogan JR, Holmboe ES, Hauer KE. Tools for direct observation and assessment of clinical skills of medical trainees: a systematic review. *JAMA* 2009;302:1316-26.
- 38 Overeem K, Faber MJ, Arah OA, Elwyn G, Lombarts KM, Wollersheim HC, et al. Doctor performance assessment in daily practise: does it help doctors or not? A systematic review. *Med Educ* 2007;41:1039-49.

Accepted: 13 September 2010