

Challenges in multisource feedback: intended and unintended outcomes

JOAN SARGEANT,¹ KAREN MANN,² DOUGLAS SINCLAIR,¹ CEES VAN DER VLEUTEN³ & JOB METSEMAKERS⁴

CONTEXT Multisource feedback (MSF) is a type of formative assessment intended to guide learning and performance change. However, in earlier research, some doctors questioned its validity and did not use it for improvement, raising questions about its consequential validity (i.e. its ability to produce intended outcomes related to learning and change). The purpose of this qualitative study was to increase understanding of the consequential validity of MSF by exploring how doctors used their feedback and the conditions influencing this use.

METHODS We used interviews with open-ended questions. We purposefully recruited volunteer participants from 2 groups of family doctors who participated in a pilot assessment of MSF: those who received high ($n = 25$) and those who received average/lower ($n = 44$) scores.

RESULTS Respondents included 12 in the higher- and 16 in the average/lower-scoring groups. Fifteen interpreted their feedback as positive (i.e. confirming current practice) and did not make changes. Thirteen interpreted feedback as negative in 1 or more domains (i.e. not confirming their practice and indicating need for change). Seven reported making changes. The most common changes were in patient and team communication; the least common were in clinical competence. Positive influences upon change

included receiving specific feedback consistent with other sources of feedback from credible reviewers who were able to observe the subjects. These reviewers were most frequently patients.

DISCUSSION Findings suggest circumstances that may contribute to low consequential validity of MSF for doctors. Implications for practice include enhancing procedural credibility by ensuring reviewers' ability to observe respective behaviours, enhancing feedback usefulness by increasing its specificity, and considering the use of more objective measures of clinical competence.

KEYWORDS humans; male; female; *feedback, psychological; clinical competence/ *standards; physicians, family, *standards/ psychology; family practice/ *standards; questionnaires; pilot projects; Nova Scotia.

Medical Education 2007; **41**: 583–591
doi:10.1111/j.1365-2923.2007.02769.x

INTRODUCTION

The purpose of formative performance assessment is to monitor and inform learners and practitioners about their performance and guide their continued learning and practice improvement. Although much has been written about assessment tools and procedures in medical education, less is known about formative outcomes in terms of learners' and practitioners' use of feedback for continued learning and improvement.¹

Although the intention is that assessment feedback will be used for improvement, the relationship between feedback and outcomes is not necessarily linear and feedback does not always achieve the desired results. Positive evaluations have been found

¹Department of Continuing Medical Education, Dalhousie University, Halifax, Nova Scotia, Canada

²Division of Medical Education, Dalhousie University, Halifax, Nova Scotia, Canada

³Department of Educational Development and Research, University of Maastricht, Maastricht, The Netherlands

⁴Department of General Practice, University of Maastricht, Maastricht, The Netherlands

Correspondence: Joan Sargeant PhD, Director of Programme Development and Evaluation, Continuing Medical Education, Dalhousie University, Clinical Research Centre, 5849 University Avenue, Halifax, Nova Scotia B3H 4H7, Canada.
Tel: 00 1 902 494 1995; Fax: 00 1 902 494 1479;
E-mail: joan.sargeant@dal.ca

Overview

What is already known on this subject

Multisource feedback (MSF) is intended to guide learning and performance improvement, but may not always result in these outcomes, especially in response to negative feedback.

What this study adds

Almost half of doctors who received negative feedback did not accept or use it. The feedback most consistently used was specific, received from patients, and addressed communication skills. The feedback least frequently used addressed clinical competence and came from medical colleagues.

Suggestions for further research

More rigorous study will inform the consequential validity of MSF, particularly to determine credible and fair assessment procedures and constructive feedback formats, and to determine the performance domains most appropriate for assessment by MSF.

to be received more positively and perceived as more accurate than negative ones,²⁻⁵ and recipients focusing on positive, not negative, aspects of their performance are more likely to improve.⁶ A meta-analysis of 600 performance appraisal feedback intervention studies showed surprising results: participants improved in a third of studies, stayed the same in a third, and actually decreased the quality of their performance in a third.⁷ In medical education, a study of the influence of medical student feedback on faculty performance showed similar findings. For teachers with low ratings, negative feedback actually led to deterioration of clinical teaching practice and lower subsequent scores.⁸

Some outcomes represent unexpected and unintended consequences of assessment. Consequential validity is a concept that links the consequences or outcomes of formative assessment, particularly its effects upon learning and improvement, with the assessment itself.⁹⁻¹¹ Assessment

feedback with high consequential validity positively influences learning and practice improvement in that it produces intended outcomes. Assessment with low consequential validity does not produce these intended outcomes and may have unintended or even detrimental consequences, such as decreased motivation, emotional distress and deteriorated performance.^{7,8} Unintended outcomes often occur due to participants' negative perceptions of the assessment process, including lack of transparency and perceived bias and lack of fairness.¹² Such perceptions can evoke negative responses to the assessment and discourage use of the feedback for learning and improvement.

Multisource feedback (MSF, or 360-degree feedback) involves a questionnaire-based process using several reviewer groups and self-assessment for formative assessment of multiple performance domains.^{13,14} It was designed for use in settings where reviewers work closely enough with those being assessed to directly observe their work and interactions with others. Its premise is that assessments are based upon observations of performance.

Multisource feedback reviewers of doctors' performance include medical colleagues, co-workers and/or patients. Although some studies have suggested that MSF is reliable and feasible for doctors and residents, others have questioned its validity.¹⁵⁻²⁰ Evans *et al.*²¹ recommended paying greater attention to the consequential and face validity of MSF to ensure that it produces desired learning and improvement outcomes.

Although earlier studies have demonstrated doctors' self-reported specific practice improvements resulting from MSF,^{17,20,22} in a recent qualitative study, we showed that some doctors responded negatively to their assessment, questioned its validity and utility and hence were disinclined to use it for improvement.²³ The purpose of the current study was to increase understanding of the consequential validity of MSF, particularly the specific use participants made of their feedback for improvements or change, and the conditions influencing this use.

METHODS

Study background

The College of Physicians and Surgeons of Nova Scotia conducted a pilot study of MSF for family doctors using the standardised Physician

Achievement Review programme, College of Physicians and Surgeons of Alberta.^{24,25} Volunteer doctor participants identified 8 medical colleague and 8 co-worker reviewers, and randomly selected 25 patient reviewers. General performance domains included communication, collegiality, professionalism, clinical performance and office management. Questionnaires included multiple domains, and items within each domain were specific for each reviewer group (medical colleagues: 31 items; co-workers: 17 items; patients: 40 items). Reviewers used a 5-point Likert scale. Doctors completed a self-assessment questionnaire that exactly mirrored the 31 medical colleague items. Participants received their compiled feedback in the form of mailed confidential reports, which provided individual and aggregate mean scores by domain and individual scores for items within each domain. An overall cumulative score was not provided. Scores above the 90th percentile were noted by a 'commendation' flag; scores below the 10th percentile and scores of '4' were noted by an 'information' flag.

Participants

Using purposeful sampling,²⁶ we identified 2 groups of doctors, those who received generally high scores from all reviewer groups and in all domains as noted by commendation flags ($n = 25$), and those who received generally average and lower scores ($n = 44$). We identified average scorers as those who received mid-range domain scores and no more than 1 flag, either commendation or information ($n = 27$), and lower scorers as those who received low domain scores and at least 1 domain information flag ($n = 17$). However, differences between the 2 groups were not always clear as most doctors in both groups received inconsistent scores and flags across domains (e.g. patient communication and clinical competence) and reviewer groups (e.g. patients and medical colleagues).

We mailed invitations to participate in an interview. Positive respondents included 12 in the higher- and 16 in the average/lower-scoring group (10 average-scoring, 6 lower-scoring). Due to research funding constraints, we conducted interviews at different times: with high scorers within 1 year of their receiving feedback reports (group A), and with average/lower scorers 2 years after their receiving feedback reports (group B).

Study design

We used interviews with open-ended and guiding questions to facilitate participants' descriptions of experiences and perceptions meaningful to them

and to explore related interpretations.²⁶ Questions explored the participants' use of their feedback results in each domain and factors influencing this.

Interviews lasted approximately 1.5 hours and were audio recorded and transcribed. The study was conducted in 2003 and 2004 by Dalhousie University Office of Continuing Medical Education.

Data analysis

We conducted the analysis as a team, using accepted analytical procedures for qualitative data.²⁶ First, using a content analysis approach, we individually reviewed and coded 2 transcripts, then discussed these and developed a coding framework. We used this framework to analyse remaining transcripts individually and met regularly to discuss emerging themes, resolve differing interpretations, and revise the coding structure as required. One researcher then compared and contrasted data within and among participants and themes, to determine and interpret relationships and confirm dominant themes; this work was guided by the research team.

RESULTS

High-scoring doctors included 10 men and 2 women, in practice for an average of 25 years. Five practised in communities larger than 50 000 and 7 in smaller communities. The 16 average/lower-scoring doctors included 12 men and 4 women, in practice for an average of 23 years. One practised in a community larger than 50 000. Compared with the Nova Scotia family doctor population, women were under-represented and rural doctors were over-represented in both groups. The average/lower-scoring group was more rural than the high-scoring group.

Notably, 15 of the 28 participants, comprising all the high-scoring group, and 3 of the average-scoring doctors reported that they interpreted their feedback as generally positive across domains (i.e. it confirmed their current practices and generally did not indicate a need for change). Hence, this group did not make changes in response to their MSF reports. The remaining 13 doctors in the combined average/lower-scoring group received feedback which they interpreted as negative in 1 or more specific domains or as generally negative across domains (i.e. as not confirming current practice and indicating need for change). Seven of these doctors who received average or lower scores reported making changes in response to their feedback (Fig. 1).

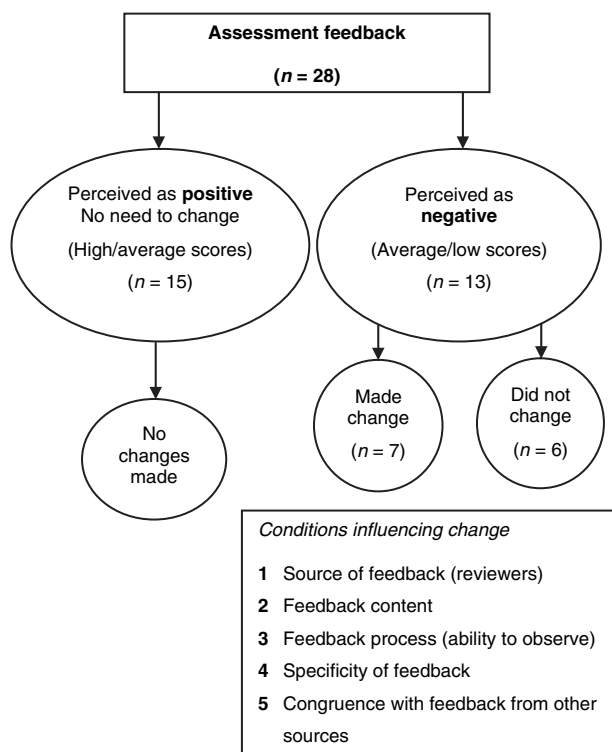


Figure 1 Influences upon decisions to change in response to multisource feedback

The following 2 sections describe the changes participants made and the influences upon their decisions to make these changes. The first section includes the perspectives of doctors who received negative feedback, and the second section includes those of doctors who received both negative and positive feedback.

Improvements or changes made

The group of doctors receiving negative feedback all reported making changes in communication with patients and/or members of the health care team. For communication with patients, they most commonly reported allowing more time for explanation and eliciting concerns:

‘I needed to communicate better about the simpler things, you know. After 15 years of practice you get used to – “There’s nothing wrong with this person so let’s move on to the next.” Now I spend a little bit more time to explain things better... just an extra minute or two can make such a difference, to listen a bit, to offer suggestions.’ (B-1)

‘I thought I explained adequately and that most of my patients would feel comfortable asking me questions if they didn’t understand, but apparently

not. Now I ask a lot of people I didn’t ask before – “Do you have any questions?” “Is there anything else going on? Anything you didn’t tell me or are afraid of?”’ (B-11)

The most frequently reported changes regarding team communication concerned improvements in written and verbal communication with pharmacists. Another doctor described how his feedback had influenced him to make broad changes in his communication with colleagues and co-workers within his small hospital:

‘Communication within this institution was poor... [As a result of this feedback] I now know everybody’s name and I have a fixed pattern that I take through the building every day – stop at the lab and the X-ray to talk to the techs, have coffee with the radiologist even if only for 3 minutes, I always see people in Emergency. I have gone out of my way to make sure that every opportunity to communicate is there.’ (B-3)

Others described participating in educational or other activities to enhance communication and professional skills. For some, these were formal learning activities, such as attending a 2-day out-of-town course at personal expense to improve patient communication skills (B-16) or record-keeping skills (B-1). Within the professional domain, 1 doctor reported seeking private consultation for stress management (B-15). Others described informal learning activities, such as reviewing student resources for enhancing patient communication skills (B-11), or observing colleagues’ interpersonal interactions at continuing medical education (CME) events (B-14). All reported positive results from their learning. Only 1 participant reported acting in response to feedback in the clinical domain (B-15). This was substantive in nature as he described conducting an audit of his patient population to determine common health problems and identify related learning needs.

These changes are significant and reflect the influence of MSF: it can be an effective tool in raising awareness and triggering action in response to an identified need. However, these results raise questions about why only some doctors changed and the types of changes they made.

Influences upon decisions to change and improve

As noted above, participants who received feedback that was generally seen as positive and confirming of

current practices reported that they did not change. For feedback interpreted as negative, responses were divided and only slightly more than half reported making any improvements. To understand this, we explored the influences upon their decisions to change or not to change. These seemed to fall into 3 broad categories:

- source (i.e. reviewer group) and content of the feedback;
- specificity of the feedback, and
- comparison of MSF scores with feedback from other sources.

Source and content of feedback

The source of the feedback appeared to influence whether or not it stimulated change.

Those making changes reported doing so in response to consistent feedback from the 3 reviewer groups (patients, medical colleagues, co-workers), or, notably, in response to patients alone. Further, 6 of 7 participants making changes attended to patient feedback preferentially over that of medical colleagues (i.e. when medical colleague feedback was less favourable than that of patients, they did not respond to it). This was also true for the 6 doctors who received negative feedback and did not make changes. They attended to their more favourable patient feedback, not to their less favourable medical colleague feedback.

To explain this, participants observed that doctors feel primarily responsible to their patients:

‘...if the patient gives you a bad report, that’s different – general practice centres around the patient.’ (B-4)

‘I was disappointed in my results until I got to the patient section, which I felt was probably the most important... they seemed quite pleased with what I was doing. But I don’t know how my medical colleagues assessed me.’ (B-2)

As the latter quote suggests, the perceived ability of medical colleagues to make the assessments also influenced responses. Participants explained that doctor colleagues rarely had the opportunity to observe them in practice:

‘When you’re asking colleagues to report on the day-to-day practice in this office, you’re asking

them to answer questions [on something] that they have no knowledge of.’ (B-13)

Although this concern pertained to behaviours assessed by both medical colleagues and co-workers, it seemed most problematic in the assessment of clinical competence, assessed only by medical colleagues. This presented a conundrum; despite agreement that doctor colleagues represented the only group with sufficient expert knowledge to make clinical assessments, they lacked opportunities to observe their family doctor colleagues’ clinical performance in their office settings:

‘It’s very difficult to have an assessment done because we’re rarely observed [by other doctors] in what we do. The people that observe us most closely are our patients and most of them are not in a position to assess aspects of clinical competency.’ (B-16)

Based upon limited opportunities for medical colleagues to observe and assess their clinical competence, participants questioned the capacity of an MSF questionnaire to adequately assess clinical competency:

‘I think that asking patients to assess the sorts of items that they’re asked is reasonable. Looking at the mechanics of running the office is reasonable. But those are all peripheral issues and the really critical issue that we all want to know is – are doctors providing competent clinical care? That’s the most difficult thing to assess and I don’t know that this tool does that.’ (B-16).

There was general consensus among participants, not just those who received negative feedback, that clinical competence should not be assessed by MSF. They suggested using a separate objective measure, such as a chart audit comparing individual practice with clinical practice guidelines.

In summary, both the source and the content or domain of the feedback influenced its acceptance. The common determining factor appeared to be perceptions of reviewers’ ability to observe and assess the particular behaviour.

Specificity

There was also consensus that some feedback was not specific enough to guide changes. Again, this appeared to be especially true of feedback from medical colleagues regarding clinical competence:

‘The part that concerned me the most was the evaluation of my clinical ability by other physicians. It was not good. So I’m saying to myself, “Okay, if that’s really me, then I’ve got to pull up my socks” – if I knew where. But I have no idea where. What’s the sense of getting an evaluation if you’re not going to be able to act on what you learned?’ (B-7)

‘The biggest problem I found with the whole approach was that there was very little in the way of useful feedback. There was a lot of subjective information but it didn’t really tell you, “How am I going to improve this?”’ (B-2)

To illustrate this point, participants critiqued examples of specific questionnaire items. Items used by medical colleagues to assess clinical competency tended to be general in nature, such as: ‘Critically assesses diagnostic information’, and ‘Selects the appropriate treatment’. By contrast, items on the patient questionnaire were more specific, for instance: ‘Your doctor clearly explained how and when to take your medications’, and ‘Your doctor adequately explained your treatment choices’. Additionally, items related to pharmacist communication were specific, for example: ‘Writes prescriptions clearly.’ Specific items were more frequently used for improvement.

It appears that, in addition to the source and content of the feedback, the degree of specificity of questionnaire items also influenced their usefulness for change. Whereas items relating to clinical competence were less specific and hence less useful, items in the communication domain were more specific and appeared more useful.

Comparison with performance feedback from other sources

Participants who received negative feedback also spoke of comparing their MSF reports with both formal and informal performance feedback received from other sources. Several participants who made changes in response to MSF explained how it confirmed other feedback, mainly that received informally from patients, medical colleagues, co-workers and/or through self-assessment. This strengthened earlier perceptions that improvement may be needed:

‘I had sloppy notes and sloppy notes are very dangerous. I knew that. This [MSF] thing made me look at my process of recording what I do... And, I did a course in record-keeping that my colleague recommended.’ (B-1)

Alternatively, doctors who reported not making changes described receiving formal performance feedback on clinical competence that was inconsistent with their negative MSF from medical colleagues. They cited this as a reason for not changing. Three of these doctors (B-4, B-8, B-13) shared with the interviewer formal performance feedback reports received from regulatory or professional bodies that recognised the high quality of their clinical practice. Understandably this caused them to question the accuracy of the clinical section of their MSF reports. In fact, 2 doctors, who were very disappointed in their feedback, sought the opinions of community colleagues regarding their ratings on clinical competence. Both reported that their colleagues disagreed with their low MSF scores.

Consistency with other feedback, both formal and informal, appeared to influence the decision to use the feedback for change.

DISCUSSION

The purpose of this study of family doctors’ use of multisource assessment feedback was to increase understanding of its use and consequential validity (i.e. the intended and unintended outcomes resulting from the feedback). The intended outcome is that MSF will be formative and that feedback, particularly negative feedback, will result in participants’ practice improvement and/or learning. In this study, positive feedback (i.e. that confirming current practice) resulted in few improvements. However, negative feedback (i.e. that suggesting a need for practice improvement) only selectively produced improvements, similar to earlier studies.^{5,7,8,23} In fact, almost half of those who received negative feedback did not accept it or use it to improve and several reported prolonged emotional distress, an unintended outcome of MSF.²⁷ This study explored the conditions explaining this unintended outcome.

Generally, the initial influence upon decisions to use feedback appeared to involve the nature of the feedback (i.e. whether it was positive or negative). Positive feedback was easily assimilated by the receiver but negative feedback led first to an appraisal of its credibility. Participants assessed credibility from several perspectives: the process of making assessments (being observed or not); source of feedback (colleagues or patients); feedback content (clinical or other); specificity of feedback (sufficient to guide change or not), and congruence with performance

feedback from other sources (matching or not matching that from other sources).

For participants who received negative feedback, credibility was most influenced by the perceived ability of reviewers to observe their performance. Participants' changed in response to feedback on a particular behaviour from reviewers who could observe their performance. For these family doctors in office practice, the reviewers who met this criterion most frequently included patients and least frequently comprised medical colleagues. Participants changed least in response to reviewers whom they perceived as unable to observe performance. In other words, perceived inability to observe contributed to the low consequential validity of the feedback.^{10,12,21} Participants recommended that reviewers should not be asked to assess what they could not observe, a recommendation consistent with the design and intent of MSF.^{13,14}

With respect to the type of improvements made, all the doctors who changed reported making changes in communication with patients. This appeared to occur for 2 reasons, because patients directly observed this behaviour and because questionnaire items for patient communication were specifically worded, providing clear direction for change. These findings are important because they show the potential for change when specific and clear feedback is given. They also suggest the strong influence that direct patient feedback can have upon doctor performance.

Conversely, feedback on clinical competence from medical colleagues was perceived as least credible and was infrequently used for improvement. Three factors appeared to contribute to this: inability to observe the performance of colleagues; lack of specificity of the items, and, for some, inconsistency with other sources of assessment feedback. There appeared to be consensus among participants that clinical competence should not be assessed using MSF. They suggested the use of assessment tools which objectively measure clinical processes and outcomes, such as chart audit.^{28,29} This is consistent with what is known about the nature of clinical competence (i.e. that it is not a generic skill but is context-specific and its assessment should also be context-specific).

Limitations of this study include the volunteer nature of participants, both in the initial pilot study and the interviews. The sensitive nature of the research may have prevented the complete sharing of negative

reactions. The role of interviewers as staff from the local academic CME office may have influenced participants to provide more socially desirable responses regarding practice change and learning following MSF. Although it was feared that the time lapse between the receipt of scores and interviews with group B participants may have affected participants' ability to remember accurately, the strength of their responses and the fact that most had their reports available for reference seemed to mitigate this effect.

These findings suggest that the consequential validity of MSF for doctors (i.e. their use of it for learning and improvement) may be low under specific conditions and suggest actions to enhance its intended outcomes. These are in keeping with the philosophical intent and procedural guidelines for MSF^{13,30} and include the following items.

- 1 Ensure the credibility of the process for doctor participants. This can be done by including reviewer groups who can observe the behaviour in question and by being particularly sensitive to behaviours which medical colleagues are able to observe. This may be more difficult for doctors who practise in isolation, whether in rural or urban settings.²³ Other ways to improve credibility are to orientate reviewers to the importance of using the 'unable to assess' questionnaire option for behaviours that are not observed, and to maintain transparency of the assessment processes.
- 2 Ensure the usefulness of the feedback by wording items as specifically as possible.
- 3 Multisource feedback programmes that do not currently include patients should consider their addition as patients do directly observe their doctors and their feedback can be powerful.
- 4 Consider assessing performance in the clinical competence domain using measures other than MSF, which can more objectively measure clinical processes and outcomes, such as chart audit.

Finally, medical practice and professional competence are multi-dimensional and composed of several domains. Different measures or tools most appropriately assess different domains. Multisource feedback can serve as a helpful tool in the box.^{1,11,31} Further research will better inform the performance domains that are most effectively assessed by MSF and those that are best assessed by other measures.

Contributors: JS substantially contributed to the conception, design, data analysis and interpretation of the

study, and drafted the paper. KM substantially contributed to the conception, design, data analysis and interpretation of the study, designed the interview questions, and revised the paper critically for important intellectual content. DS, CvdV and JM substantially contributed to the design, data analysis and interpretation of the study, and revised the paper critically for important intellectual content. All authors approved the final version of the paper.

Acknowledgements: the authors acknowledge Suzanne Ferrier MA, research associate, Dalhousie University Office of CME, for her assistance with data collection and analysis, and Philip Muirhead MD, family doctor, Musquodoboit Harbour, Nova Scotia for his assistance with data analysis and for ensuring this work was grounded within the context of family practice.

Funding: this study was funded by the Nova Scotia Health Research Foundation and the Medical Council of Canada.

Conflicts of interest: none.

Ethical approval: this study was approved by the Research Ethics Board of Dalhousie University.

REFERENCES

- van der Vleuten CPM. The assessment of professional competence: theoretical developments, research and practical implications. *Adv Health Sci Educ Theory Pract* 1996;**1**:41–67.
- Halperin K, Snyder CR, Shenkel RJ, Houston BK. Effects of source status and message favorability on acceptance of personality feedback. *J Appl Psychol* 1976;**61**:142–7.
- Morran DK, Stockton RA. Effect of self-concept on group member reception of positive and negative feedback. *J Counsel Psychol* 1980;**27** (3):260–7.
- Stone D, Stone E. The effects of feedback consistency and feedback favorability on self-perceived task competence and perceived feedback accuracy. *Organizational Behav Hum Dec Processes* 1985;**36**:167–85.
- Kernis M, Johnson E. Current and typical self-appraisals: differential responsiveness to evaluative feedback and implications for emotions. *J Res Personal* 1990;**24**:241–57.
- McFarland C, Miller DT. The framing of relative performance feedback: seeing the glass as half empty or half full. *J Pers Soc Psychol* 1994;**66**:1061–73.
- Kluger AN, DeNisi A. The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychol Bull* 1996;**119**:254–84.
- Litzelman DK, Stratos GA, Marriott DJ, Lazaridis EN, Skeff KM. Beneficial and harmful effects of augmented feedback on physicians' clinical teaching performance. *Acad Med* 1998;**73**:324–32.
- Boud D. *Enhancing Learning through Self-assessment*. London: Kogan Page 1995.
- Gielen S, Dochy F, Dierick S. Evaluating the consequential validity of new modes of assessment: the influence of assessment on learning including pre-, post- and true assessment effects. In: Segers M, Dochy F, Cascallar E, eds. *Optimising New Modes of Assessment. In Search of Qualities and Standards*. Dordrecht: Kluwer Academic Publishers 2003.
- Van der Vleuten CPM, Schuwirth LWT. Assessing professional competence: from methods to programmes. *Med Educ* 2005;**39**:309–17.
- Messick S. Validity. In: Linn RL, ed. *Educational Measurement*. New York: MacMillan 1989.
- Bracken DW, Timmreck CW, Church AH. *The Handbook for Multisource Feedback*. San Francisco: Jossey-Bass 2001.
- Tornow WW, London M. *Maximizing the Value of 360-degree Feedback*. 1st edn. San Francisco: Jossey-Bass 1998.
- Hall W, Violato C, Lewkonian R *et al*. Assessment of physician performance in Alberta: the physician achievement review. *CMAJ* 1999;**161**:52–7.
- Violato C, Lockyer J, Fidler H. Multisource feedback: a method of assessing surgical practice. *BMJ* 2003;**326** (7388):546–8.
- Lipner RS, Blank LL, Leas BF, Fortna GS. The value of patient and peer ratings in recertification. *Acad Med* 2002;**77** (Suppl 10):64–6.
- Sargeant J, Mann KV, Ferrier SN, Langille D, Muirhead PD, Sinclair DE. Responses of rural family physicians and their colleagues and co-worker raters to a multisource feedback process: a pilot project. *Acad Med* 2003;**77** (Suppl 10):542–4.
- Archer JC, Norcini J, Davies HA. Use of SPRAT for peer review of paediatricians in training. *BMJ* 2005;**330** (7502):1251–3.
- Bennett H, Gatrell J, Packham R. Medical appraisal: collecting evidence of performance through 360-degree feedback. *Clin Manag* 2004;**12**:165–71.
- Evans R, Elwyn G, Edwards A. Review of instruments for peer assessment of physicians. *BMJ* 2004;**328** (7450):1240.
- Fidler H, Lockyer JM, Toews J, Violato C. Changing physician practices: the effect of individual feedback. *Acad Med* 1999;**74** (702):14.
- Sargeant JM, Mann KV, Ferrier S. Exploring family physicians' reactions to multisource feedback: perceptions of credibility and usefulness. *Med Educ* 2005;**39**:497–504.
- College of Physicians and Surgeons of Alberta. Physician Achievement Review (PAR). <http://www.par-program.org/>. [Accessed 2 March 2006.]
- College of Physicians and Surgeons of Nova Scotia. Nova Scotia Physician Achievement Review (NSPAR) Program. <http://www.nspar.ca/>. [Accessed 2 March 2006.]
- Rice PL, Ezzy D. *Qualitative Research Methods*. South Melbourne, Victoria: Oxford University Press 2000.
- Sargeant J, Mann K, Sinclair D, van der Vleuten C, Metsemakers J. Understanding the influence of emotions and reflection upon multi-source feedback acceptance and use. *Adv Health Sci Educ* 2006; DOI: 10.1007/s10459-006-9039-x.

- 28 Donen N. Mandatory practice self-appraisal: moving towards outcomes-based continuing medical education. *J Eval Clin Pract* 1999;**5**:297–303.
- 29 Piterman L, Nelson M. Clinical audit – linking continuing medical education (CME) and practice assessment (PA). *Aust Fam Physician* 1997;**26** (Suppl 2):91–5.
- 30 Gray A, Lewis A, Fletcher C, Burke E, Mackay J, Kubelius E, Lindley P. 360-Degree Feedback. Best Practice Guidelines. <http://www.psychtesting.org.uk/search.asp?sub=true>. [Accessed 2 March 2006.]
- 31 Accreditation Council of Graduate Medical Education. ACGME Outcome Project. <http://www.acgme.org/Outcome/>. [Accessed 2 March 2006.]

Received 21 March 2006; editorial comments to authors 13 July 2006, 10 October 2006; accepted for publication 31 October 2006