Development of a multisource feedback instrument for clinical supervisors in postgraduate medical training

Mayen Egbe and Paul Baker

Abstract – Medical training is a complex endeavour and analysing its quality is not a simple task. The accuracy of information, particularly of data gathered from trainees, will depend greatly on its source, on perceptions relating to confidentiality and on the uses to which the data are put. These factors should guide our choice of feedback instrument. In addition, in times of 'survey fatigue', we have a duty to be efficient. This paper discusses the piloting of the 'feedback on performance for trainers' tool for clinical supervisors of geriatric medicine trainees in the North Western Deanery. This tool is a multisource feedback instrument that can be used to gather information specifically on the perceived quality of clinical supervision. The tool's design takes into account opinions relating to confidentiality and content validity that have been expressed by trainees, trainers, educationalists and administrators. The tool is relatively easy and quick to use, taking about 10 min of trainee or trainer time. Administrative support is needed but the workload should not be onerous, especially if an on-line process can be developed. Strong evidence to support the validity of this instrument has been collected. The next step is the development and evaluation of the approach as an online process.

Introduction

It is important that we understand the quality of medical training in our geriatric medicine training locations. In assessing what goes on, there is no definitive 'truth' but different opinions that must be interpreted. In this context, 'triangulation' is the term usually applied to describe the cross-checking of information. There must be an acceptance that one piece of evidence is just that, nothing more than uncorroborated opinion.

Many different skills and attributes of clinical supervisors are relevant to the discussion. Some, such as training qualifications and continuing professional development (CPD), will be matters of record that need not be assessed by colleagues. Other attributes are more subjective and must be assessed by feedback from colleagues; these include overall competence, standards of supervision, quality of feedback provision, responsiveness and accessibility.

Mayen Egbe, consultant physician, Royal Bolton Hospital; **Paul Baker,** consultant physician, Royal Bolton Hospital and head of foundation school, North Western Deanery, Manchester

With this in mind, a 360º appraisal that focuses on educational aspects of a trainer's work can be envisaged. As with the mini-peer assessment tool (PAT) used in foundation training, peer review and self-assessment using a shared assessment tool would be part of this process.

Assessments can be performed manually using paper returns, provided a mutually acceptable 'honest broker' can be found to collate the information and feed it back confidentially. The North Western Foundation School intends to use the programming to gather information from trainees into its e-portfolio to generate data from trainers too, thereby beginning to gather evidence that could determine the validity[1] of their multisource feedback approach, which we have little or no data on at present.

Choice and design of assessment tools

It is possible to design assessment tools that are very user-friendly but which collect information that is so brief as to be meaningless. Alternatively, a tool could be made to gather extremely thorough and precise data but that takes so long complete that nobody will provide all of the required data. The design of all assessments is therefore a compromise.

The design of all assessments should pay attention to brevity (probably no more than 20 questions will be well tolerated), clarity of questions and so on.[2] Content validity will be improved by gathering information only on the attributes that are being measured. Four-point scales have been used to avoid 'fence-sitting' though a 'not applicable' option (or simply omitting an answer) will often be appropriate.

General principles

'Survey fatigue' is a real issue, so data collection needs to be as efficient as possible.

We have gathered opinions from trainees, trainers and education staff about how best to collect accurate and honest reports in this sensitive area. These consultations highlighted several issues relating to the collection of feedback from junior doctors.

- Trainees have very real fears about the confidentiality of feedback that is related to individual trainers.
- In giving negative feedback on trainers, trainees might feel ungrateful, disloyal or endangered.
- Trainees are less hesitant to give frank feedback about units or institutions.

- The source of the query, coupled with the anonymity and perceived use of the feedback, might influence the responses.
- Many junior doctors are not well versed in education theory relating to curricula or learning environments.
- There is inherent conflict of interest in giving frank feedback relating to standards of their own training, especially for activities around e-portfolios.
- Tools such as the mini-PAT and other online 360º feedback methods are now familiar to even junior trainees.
- Poor response rates will hinder any feedback tool that isn't 'compulsory'.
- 'Survey fatigue' is a real issue.

For most trainees, providing feedback about their boss is a very sensitive area. There are real fears of confidentiality and it seems likely that many trainees will never give uncomplimentary feedback unless anonymity is assured. Conventional surveys are the least likely to be answered frankly, especially those that are web-based or linked to the responder's portfolio. However, anonymous paper questionnaires seem to be more workable. The North Western Foundation's learning portfolio has ensured that 360º appraisal is now well-accepted amongst most junior doctors who have seen it in operation. They trust, from personal experience, that the appraisee need never know which appraisers gave a particular answer.

Trainer attributes

We have already alluded to what attributes might be considered in a trainer's performance: supervision, ability to provide feedback, responsiveness to feedback, accessibility and so on. The concept of 'good training' is an elusive one and we usually describe the attributes of good trainers rather than their performance. There have been many descriptions of what makes a good trainer[3,4] and these attributes need to be stressed in any 360º appraisal; non-education aspects of a senior's performance will be covered elsewhere. The Academy of Medical Educators has recently defined training standards in great detail.[5] Common themes from review of all available information sources include assessments of general credibility as clinician and educator, background knowledge, interest in training and trainees, supervision style and technical training ability.

The 360º appraisal is an assessment tool used in clinical practice to assess the attributes of a clinician's role across a number of domains that would be otherwise difficult to evaluate with conventional measurement tools.

As medical educators charged with training tomorrow's specialists, it is essential that clinical supervisors perform to high standards in all aspects of trainee supervision. There has been, however, no robust and validated assessment tool that specifically targets a clinician's educational role as a supervisor of trainees in the workplace. The 'feedback on performance for trainers' 360º appraisal tool has been developed by the North Western Deanery for this purpose. It addresses the different aspects of clinical supervisors' role and has potential as an important tool for gathering and giving feedback to clinical

supervisors. Following a format similar to those of existing assessments in the clinical setting, feedback will be sought from colleagues, co-workers, trainees and the assessed individual.

Literature review

Multisource feedback is widely used in the business world. It has gained acceptance in the medical world as the need for assessment and maintenance of competence of doctors continues to receive worldwide attention.[6,7] In recent times, thinking about competence has shifted. Medical expertise and clinical decision-making

---

Box 1. Questions asked during 360 degree appraisal of trainer performance.

A. General credibility as clinician and educator
- Clinical credibility
- Patient-centred practice
- Professionalism, integrity and ethics
- Organisation, good management of time and resources
- Self-awareness

B. Background education knowledge
- Knowledge of curriculum
- Theoretical education knowledge
- Evidence-based training practice
- Critical analysis, research awareness
- Governance, wider education environment, regulation

C. Interest in training and trainee
- Learner-centred, operates with regard to equality/diversity
- Interested in training and trainees, is accessible/available
- Responsive to trainee and needs
- Pays attention to training environment
- Commitment to education, motivation

D. Supervision and training style
- Supervises but allows controlled risk
- Supportive to the learner
- Challenging when necessary
- Consistent in approach

E. Technical training ability
- Performs on the job/workplace-based learning
- Good at feedback/communication
- Demonstrates/teaches reflection
- Performs training needs assessment
- Awareness and use of assessment methods
- Design and planning learning activities

Assessors were informed that this was a pilot relating to quality control and quality management of training and asked to grade five aspects of being a trainer, A–E, as applicable to the appraisee, using a 4-point scale (4: really good, great, highly commended; 3: fine, useful, good enough; 2: below the mark, patchy, needs to be better; 1: dire, very poor, badly done if at all.) Assessors were informed that a score of 4 or 3 was deemed 'good enough', while a score of 1 or 2 was not. If they had not witnessed this aspect and therefore could not form an opinion, they were instructed to indicate 'N/S'. Space was also provided for free comments. The form concluded with a space for the assessor to comment on the questionnaire itself, including how long it took to complete, the layout and the content.

---

are increasingly recognised only as components of competence. Now, communication skills, interpersonal skills, collegiality and professionalism are also considered when assessing doctors.[8] The profession has identified and set out professional standards for practice.[5,6] The General Medical Council (GMC) has set out the attributes it expects of a doctor across seven domains,[9] and the domains about which information is collected in multisource feedback assessments mirror these. Specialties within the profession have taken things a step further by developing specialty-specific Multisource feedback assessments with domains that focus on the core attributes of a practitioner within the specialty.[10–12]

There is, however, little information in the medical literature about the existence of a robust tool that can assess the different facets of a medical educator's responsibilities. Multisource feedback has been used to evaluate the teaching and professionalism[13] and professional skills of clinical directors,[14] but never to assess the role of the clinical supervisor in the context of medical education.

Methods

Validating the questionnaire

In piloting the questionnaire, we needed to show that the data collected was valid by showing that (a) it measures what it sets out to measure, (b) it is reliable, (c) it is accepted by the target population, and (d) it is consistent, eliciting responses similar to those gained by trainees in another location. Finally we needed to prove that the information gathered was of use/value (ie the assessment's utility).

Participant seniors were asked to complete the survey document A (Quality Management Pilot 1) either in paper form or online and to tick the box 'I am the trainer, ie this is my self-assessment'. The form was returned to an independent administrative team at the Education Centre. In line with previous research, each consultant was asked to provide the names of at least 12 colleagues or trainees to act as appraisers.[15–18] The proportions were at the trainer's discretion.

Data was analysed for descriptive statistics, comparison of both groups, agreement studies, validity, acceptability and feasibility using Microsoft Excel 2007 and Statsdirect version 2. A statistician was consulted for statistical advice.

Individual geriatrics programmes were uncomfortable with piloting the assessment tool but two interested Trust Directors of Medical Education were enrolled, in Bolton and Morecambe Bay. These directors had themselves been considering how to obtain structured trainer feedback. The 360º appraisal form is shown in Box 1.
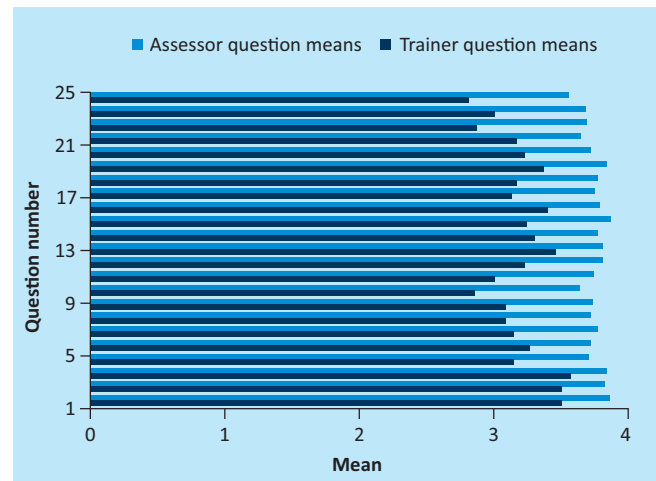

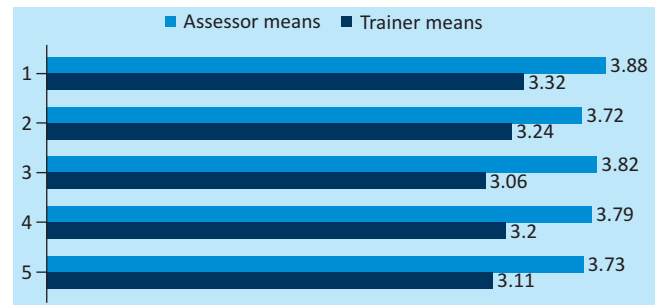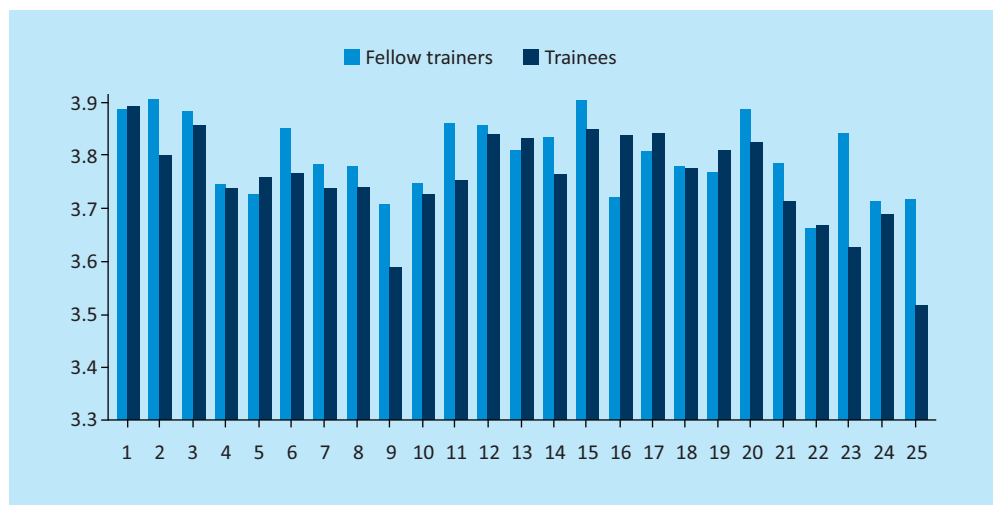Fig 1. Response means to 360 degree appraisal for all 25 questions.

Fig 2. Response means to 360º appraisal for all five domains.

Fig 3. Comparison of trainees and fellow trainers as assessors for each of 25 'questions'.
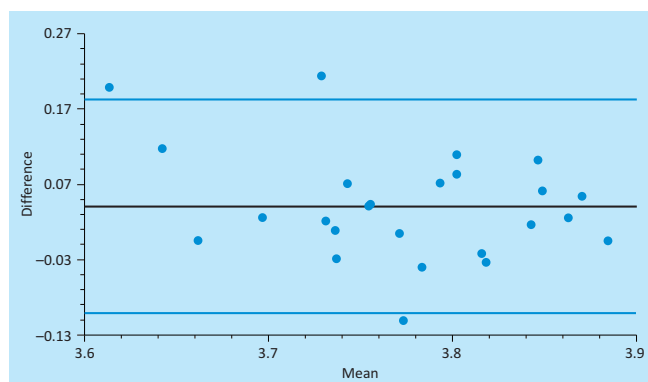
Fig 4. Agreement plot for trainee and fellow trainer assessment scores. The plot shows a random scatter of scores between +2 and −2 standard deviations (blue lines) within which 95% of scores lie. These limits of agreement suggest that the difference between trainee and fellow trainer scores lies within 0.18 points. This is a small difference on our measurement scale, so we can be confident that 95% of the time, the two assessor groups (trainees and fellow trainers) will give practically the same answer (where 'practically the same' means 'within 0.18 points').

Results

Quantitative feedback

Peer assessments were completed for all of the 31 trainers who participated in the pilot. A total of 243 assessors, 128 trainees and 115 fellow trainers or other colleagues, also took part.

The mean time taken to complete the survey was 9.5 min for assessors and 6.8 min for trainers.

Mean scores for all questions are illustrated graphically in Fig 1. Fig 2 illustrates mean scores for domains 1–5. There were no outlying trainers. There is a consistent finding that trainers underrate themselves.

Comparison of trainees and fellow trainer/colleague scores are shown in Fig 3. Fig 4 shows an agreement plot between trainee and fellow trainer scores suggesting good agreement between the two groups. Furthermore, the independent t-test suggests no significant differences between fellow trainer and trainee scores (p-value (two-sided)=0.0762 (statistical significance threshold of 0.05)).These results suggest that these two groups do not necessarily have to be surveyed separately as their scores do not significantly differ.

Qualitative feedback

Feedback about the questions asked, the layout of the questionnaire and the feasibility and utility of the process was obtained from trainers in the form of comments in free text boxes. Selected comments are reproduced in the paragraphs that follow. Free text comments were scanty compared to the other survey results but some themes were evident.

Online administration

By far the biggest single group of comments related to method of administration. The pilot 360º tool depended on e-mailed, or even paper, forms but many alluded to the potential advantages of web-based methods.

'Perhaps more use of drop down menus could improve ease of completion.'

'If it is intended that the questionnaire be completed electronically, it should be custom-made for viewing on screen.'

'Online survey might be easier.' (several similar remarks)

'Have you thought about using *Survey Monkey*?'

Number of assessors

The average number of assessors was approximately eight per trainer, less than the 12 suggested after a survey of the established literature. In foundation training, mini-PAT feedback has been given with as few as four assessors. The assessment of training by other colleagues as well as by immediate trainees may need to be emphasised further.

'It is difficult to think of 12 trainees currently with me. Would any fewer do?'

Suggestions for improvement

Positive feedback, such as 'well laid out', 'easy to navigate', 'good' and related comments, by far outweighed negative feedback.

'This is something that could be introduced as part of the routine appraisal of educators.'

Aside of some non-specific criticism, such as 'could be better' and 'editing could be easier', there were suggestions from assessors, such as those requesting more explanation of the terms used and more opportunity for feedback.

'Would prefer more opportunity to comment in 'free text' under each A to E heading to elaborate further and be more specific in some areas.'

'OK [ie the questionnaire], not sure that the questions and scoring system will allow sufficient differentiation between excellent, good and moderate trainers when there are polite people filling in the form.'

'Marking is difficult as 4 is good, great etc and 3 is good enough, needs more options.'

'Why don't you include a question such as, "would you be happy for this doctor to treat your family?"'

The discussion about descriptors rather than scores is topical in the light of national deliberations about workplace-based assessment in general, and in foundation training in particular. There seems a general movement away from scores towards descriptors, which would be in keeping with some of the suggestions made in this pilot study.

Discussion

A large amount of data was collected, of which only a small proportion is used here to illustrate the results.

The 360º appraisal for trainers proved difficult to pilot because of the sensitivities involved in personal feedback and the level of administrative support needed to administer it. The sensitivities were overcome by maintaining confidentiality of responses and

keeping the identity of assessors anonymous. The presence of an administrative assistant who was perceived as neutral with no competing interests was advantageous. Feedback was given either by the Director of Medical Education or sent directly to the trainer if so preferred. The data were not publicly accessible.

A consistent finding is that we underappreciate our own abilities, a feature also seen in foundation mini-PAT assessments. Few trainers would have had sufficient trainees of specialty registrar (StR) or specialist registrar (SpR) grades in the recent past to get truly pooled or anonymous feedback. Most authorities recommend a minimum of 11 assessors for truly representative multisource feedback, but as few as four have been used in some settings. Trainers were therefore encouraged to include trainees of any grade and to utilise senior colleagues, fellow trainers and professions allied to medicine – anyone, in fact, who has witnessed the appraisee in action in a training situation. On average we managed eight non-trainee assessors for each trainer and a comparison of their scores with those of the trainees is shown in Fig 3. The inclusion of feedback from both colleagues and trainees provides good evidence for the validity of this approach – whatever the training is, trainees and colleagues are seeing the same phenomenon.

Anonymity is vital for honest 360º feedback. This requires an 'honest broker' whose objectivity is trusted by the appraisee and whose ability to maintain confidentiality is trusted by the appraisers. The administrative support needed is significant. The introduction of an online process would address these problems and would acknowledge the qualitative feedback on ease of completion.

If the right questions were asked and links made between trainees and trainers, it would theoretically be possible to achieve the same granularity of information provided by this survey from the National Trainer Survey when all trainers are registered with the GMC. That endeavour could, however, be fraught with considerable administrative and logistical challenges.

Conclusions

The techniques outlined in this paper have been designed with logic and efficiency in mind. Their use has been shown to be practical and quick. Evidence to support the tools' validity has been gathered, but its reliability is less certain as yet.

Some conclusions can be drawn from the information gathered.

- The assessment processes were generally well accepted by trainees and trainers.
- The tool is simple and quick to use.
- Self-assessment by trainers is often less favourable than others' opinions.
- More descriptors and guidance on gradings would be appreciated.
- Useful suggestions for refining the design and administration were received.

- Finding sufficient assessors for a 360º appraisal can be challenging.
- Many issues with 360º appraisal could be solved by an online process.

Acknowledgements

References

1 Downing SM. Validity; on the meaningful interpretation of assessment data. *Med Educ* 2003;37:830–7.
2 Williams A. How to … write and analyse a questionnaire. *J Orthod* 2003;30:245–52.
3 Hayden J. William Pickles Lecture. Young ambition's ladder. *Br J Gen Pract* 2003;53:143–8.
4 Wright S, Kern DE, Kolodner K *et al*. Attributes of excellent attending-physician role models. *New Engl J Med* 1998;339:1986–93.
5 Academy of Medical Educators. *Professional Standards*. London: Academy of Medical Educators, 2009.
6 Southgate L, Hays RB, Norcini J *et al*. Setting performance standards for medical   practice; a theoretical framework. *Med Educ* 2001;35: 474–81.
7   Epstein RM, Hundert EM. Defining and assessing professional competence. *JAMA* 2002;287:226–35.
8 Violato C, Lockyer JM, Fidler H. Multisource feedback: a method of assessing surgical practice. *BMJ* 2003;326:546–8.
9 Anon. Medical Practice. Guidance for Doctors. London: General Medical Council, 2006.
10 Violato C, Lockyer JM, Fidler H. A multisource feedback program for anesthesiologists. *Can J Anaesth* 2006;53:33–39.
11 Violato C, Lockyer JM, Fidler H. Assessment of psychiatrists in practice through multisource feedback. *Can J Psychiatry* 2008;53:525–33.
12 Lockyer JM, Violato C, Fidler H, Alakija P. The assessment of pathologists/laboratory medicine physicians through a multisource feedback tool. *Arch Pathol Lab Med* 2009;133: 1301–8.
13 Berk RA . Using the 360 degrees multisource feedback model to evaluate teaching and professionalism. *Med Teach* 2009;31:1073–80.
14 Palmer R, Rayner H, Wall D *et al*. Multisource feedback: 360-degree assessment of professional skills of clinical directors. *Health Serv Manage Res* 2007 20;3:183–8.
15 Violato C, Lockyer JM, Fidler H. Assessment of radiology physicians by a regulatory authority. *Radiology* 2008;247:771–8.
16 Ramsey PG, Wenrich MD, Carline JD *et al*. Use of peer ratings to evaluate physician performance. *JAMA* 1993;269:1655–60.
17 Violato C, Lockyer JM, Fidler H. Assessment of paediatricians by a regulatory authority. *Pediatrics* 2006;117:796–802.
18   Lockyer JM, Violato C, Fidler H. The assessment of emergency physicians by a regulatory authority. *Acad Emerg Med* 2006;13:1296–303.

Address for correspondence: Dr Mayen Egbe, Royal Bolton Hospital Foundation Trust, Minerva Road, Bolton, Lancashire, BL4 0JR. Email: Mayen.Egbe@boltonft.nhs.uk