



A critical analysis of mini peer assessment tool (mini-PAT)

Aza Abdulla

Consultant Physician, Princess Royal University Hospital, Bromley Hospitals NHS Trust, Farnborough Common, Orpington, Kent BR6 8ND, UK. E-mail: aza.abdulla@bromleyhospitals.nhs.uk

DECLARATIONS

Competing interests

AA is a College tutor for the Royal College of Physicians (UK) and a member of the British Geriatric Society's clinical effectiveness and practice committee

Funding

Not applicable

Ethical approval

Not applicable

Guarantor

AA

Contributorship

AA was the sole contributor

Acknowledgement

The author is grateful to Dr S Wilkinson for his recommendations and review of the manuscript

The structured evaluation of doctors' performance through peer review is a relatively new phenomenon brought about by public demand for accountability to patients. Medical knowledge (as assessed by examination score) is no longer a good predictor of individual performance, humanistic qualities and communication skills. The process of peer review (or multi-source assessment) was developed over the last two decades in the USA and has started to pick up momentum in the UK through the introduction of Modernizing Medical Careers. However the concept is not new. Driven by market forces, it was initially developed by industrial organizations to improve leadership qualities with a view to increasing productivity through positive behaviour change and self-awareness.

Multi-source feedback is not without its problems and may not always produce its desired outcomes. In this article we review the evidence for peer review and critically discuss the current process of mini peer assessment tool (mini-PAT) as the assessment tool for peer review employed in UK.

Introduction

The evaluation of doctors' performance is a relatively new phenomenon. It was brought about by demands for public accountability of the profession's performance, and the need to ensure the safety of the public following serious concerns about poor performance by some doctors. In the UK, the General Medical Council (GMC) – the legal body whose function under the medical act is 'to promote, protect and maintain the health and safety of the public by ensuring proper standards...' – produced its first report on performance procedures only in 1997. The report stated that in addition to tests of competence to assess knowledge and basic skills, the performance of doctors who enter the procedures should also be assessed through peer review of practice in the workplace¹. However the process of assessing performance through seeking the views of others has been slow to pick up, probably because the medical profession has been geared towards tangible scientific evidence and quantitative research. Personal view(s) has been historically devalued and its validity questioned merely because it represents the unsubstantiated opinions of others, but

also because it probably challenges the long held concept that 'doctor knows best'.

The evidence for peer assessment

Peer assessment, multi-source assessment, multi-source feedback, 360 degree feedback, 360 degree appraisal, peer review, and peer rating are different names given to essentially the same process whereby the individual receives formal feedback on his/her performance at work from peers, subordinates and superior managers.

The concept was initially developed by industrial organizations and companies in the western world, largely driven by market forces. It was originally devised to improve leadership qualities, since managers and directors within an organization received little feedback on their performance, thereby limiting their learning opportunities. There was also a desire to move away from the top-down single-person approach to an individual's performance, to a fairer and perhaps more accurate system which offers a more rounded or multi-faceted overview (multi-source or 360 degree feedback). In the mid-eighties only 10% of US companies were using these systems² but in

recent years there has been a rapid increase in the popularity of multi-rater feedback, even among a wide range of public sectors in the UK. The technique should offer individuals insight into the way others perceive their performance based on their workplace behaviour and provide an opportunity to reflect on one's conduct. Several studies have shown improvement in overall performance following 360 degree feedback through increased motivation among staff. This in turn translated into increased productivity, and also brought about positive behaviour change and increased self-awareness, seen as fundamental for the progress of any organization. However, an improved outcome is by no means the rule. One study demonstrated only 50% improvement in performance of the supervisors who received 360 degree feedback.³ An earlier paper⁴ showed that in a third of cases feedback resulted in decreased performance. More recently Bret and Atwater⁵ showed that individuals receiving negative feedback may be discouraged and even react with anger. These authors suggested that this may be related to the manner in which the process is implemented and the way feedback is provided rather than a flaw in the concept itself.

It is important to note that multi-source feedback conducted in large organizations is purpose-designed by professional companies commissioned to undertake the process. These companies employ groups of psychologists and use a variety of psychometric tests. Many of them quote high content validity and reliability (correlation coefficient in the range of 0.7-0.8 for tools assessing communication skills, working relationships, team development and stress-coping strategies).

The medical profession has been slow to adopt these techniques of performance assessment. The medical system in the USA, by the very nature of its practice, acknowledged this need earlier. They accepted that board examinations do not provide information about areas of performance such as interpersonal skills and communication, and that a new evaluation tool needed to be sought. Measures to assess clinical competence and performance were therefore developed in the early eighties and subsequently refined^{6,7} and extended to include medical education.^{8,9} Initially peer assessment through the use of global performance ratings was introduced as an evaluation mechanism for recertification of practicing physicians by specialty boards including the American Board of Internal Medicine,¹⁰ and in 1993, Ramsey *et al.* published the first study to assess the feasibility

and measurement characteristics of peer rating.¹¹ This study validated the view that it is feasible to obtain assessments from professional colleagues in areas of clinical practice, humanistic qualities and communication skills.

Ramsey's study was limited to physicians practicing internal medicine, but subsequent studies from other disciplines validated the reliability of multi-source feedback and peer rating in other subspecialties including surgery,¹² obstetrics and gynaecology,¹³ and intensive care.¹⁴ Although validity and reliability was demonstrated, it is important to note that each study developed its own assessment questionnaire with the parameters under evaluation varying from as few as ten¹³ to as many as 34.¹² Similarly the number of raters providing feedback in these studies varied widely. Nonetheless these studies demonstrated the reliability of multi-source questionnaires across different settings.

The paper by Ramsey *et al.*¹¹ is considered a landmark study not only because it was a forerunner for other studies but also because of its important conclusions. It demonstrated that ratings from eleven peer physicians are needed to provide reliable assessment and that neither the method of selection of assessors or the relationship between the person being rated and the rater substantially affected the results. It also showed that although a strong correlation coefficient (in the range of 0.5 to 0.6) was seen between peer ratings of medical knowledge and American Board Examination scores, there was a low correlation (<0.15) between ratings of humanistic qualities and examination scores. The implication is that medical knowledge (as assessed by examination score) was a poor predictor of communication skills, and interpersonal relationship.

The principle of multi-source feedback and its effectiveness is therefore supported by considerable research. It has certain strengths and provides the individual being assessed with an overview of how others see him/her at work. It also offers an opportunity to compare 'self perception' with peer perception and allows comparison with the average peer group. The tool should also identify strengths and weaknesses, and highlights those areas which need to be worked on, through producing an agreed-action development plan following formal feedback.

It is, however, important to note that raters are offering their overall 'perception' on the quality of performance of a colleague rather than assessing a structured task (e.g. performing a physical examination). As such it is a personal view which is not

only affected by the standards of the particular rater but the quality of the results may also be influenced by personal relationships, stakes and equivalence.¹⁵⁻¹⁷

In the UK the only example of a validated peer review assessment tool is the Sheffield Peer Review Assessment Tool (SPRAT).¹⁸ The questionnaire was originally designed as a voluntary appraisal tool for paediatric consultants to assess components of performance as described by the GMC and the Royal College of Paediatrics and Child Health. The tool was field tested and later modified following feedback received from volunteers and psychometric evaluation to contain 24 questions covering the five domains of good medical practice. Ratings were given on a six point scale, whereby '1' was equal to 'very poor' and '6' was equal to 'very good'. '4' was 'satisfactory' and considered to be the pass mark. A box for free text comments and observations was also provided at the end of the questionnaire. As the questionnaire was mapped directly to the standards of good medical practice as defined by the GMC,¹⁹ content validity had therefore been established.

SPRAT has been used in the South Yorkshire and South Humberside Deanery to assess paediatricians in training and to demonstrate the reliability of the tool.²⁰ Of the 112 doctors who were assessed, 93 (83%) scored an overall mean of 4.5 or more (4 = pass). With a 95% confidence interval of ± 0.5 , on the basis of generalizability theory it was concluded that only four raters were sufficient to make a reasonably confident decision of satisfactory competence or that doctor was in difficulty.¹⁸ The mean time to complete the questionnaire by the rater was 6 minutes, confirming its feasibility.

With the implementation of Modernizing Medical Careers and introduction of the Foundation programme as the forerunner of the process of reforming the structure of training for doctors, the Postgraduate Medical Education Training Board (PMETB) embarked on formulating an integrated set of assessment tools for postgraduate training.^{21,22} The program acknowledged that medical practice should not only focus on scientific knowledge but that other parameters such as communication skills, team work, and humanistic qualities have important effects on patient care and should be taken into consideration. It highlighted the importance of feedback from peers by a formal assessment tool that showed uniformity thereby allowing comparison against peers in the same stage of training. The introduction of a formative assessment tool, called mini-PAT (Peer Assessment Tool)

was therefore incorporated as a requirement for good medical practice.

The mini-PAT is a shortened version of SPRAT consisting of 16 questions only but with a similar global scoring system and space for free text. The procedure follows well defined guidelines. The trainee is asked to nominate and provide contact details of eight people who will act as assessors (raters). They should be healthcare professionals rather than administrative staff or patients. They maybe supervising consultants, GP principals, staff grades, specialist registrars, senior house officers, other foundation doctors, nurses, or professionals allied to medicine. The trainee is reminded to choose raters from a variety of professional backgrounds and from the different clinical environments the trainee works in. Consent for rater recruitment is a verbal process where the doctor being assessed approaches the rater to agree to participate in the exercise. The trainee should also complete a self-assessment using the same questionnaire.

The list of raters is sent back to central office and assessment forms are sent directly to the nominated assessors to ensure that the views of the individual assessors remain anonymous to the trainee. The raters are also sent an explanatory letter informing them of the process. The responses received from the raters are collated at the national centre in Sheffield and fed back to the educational supervisor. These results are formatted to compare self-rating to the mean score from the assessors and also the average rating of peer group at the same stage of training.

Criticisms and suggestions for modification

Validity and reliability

It is no secret that mini-PAT lacks sufficient field evaluation and has not gone through any stringent criteria that are required for the validation of an assessment tool. No evidence of reliability has been published to date and in fact even the parent version of mini-PAT, SPRAT, has only been formally validated through one field study.¹⁸ Nonetheless since the questionnaire conforms to good medical practice as defined by GMC,²² its content validity will not be questioned. Similarly as it is a shortened version of SPRAT (which takes 6 minutes to complete), there should not be a problem with its feasibility. However we are unaware of the criteria used to reduce the questionnaire by 30% (16 in mini-PAT of the 24 questions used in

SPRAT). Furthermore, although mini-PAT requests eight raters to provide feedback, no reference has been made as to the minimum number of raters needed to produce a valid result. This creates a potential problem; with data from a small number of raters, it is conceivable that the simple statistical methods used (like the mean) can produce erroneous results which may not reflect the trainee's actual performance. Although this has not been shown by the SPRAT data,¹⁸ it is important to note that this study was small, comprising only 112 trainees. Using a larger number of raters should reduce the measurement error considerably. Alternatively employing different statistical methods designed for small sample size may be more informative.

Choice of raters

According to Ramsey, the choice of raters as identified by the trainee is appropriate and does not appear to significantly affect the reliability of the overall results. However, there remains a concern that the trainee may not select the right mix of raters. In medical students it was noted that those with low levels of peer assessed interpersonal attributes were more likely to select other low-rated classmates to rate and be rated by.²³ It is also known that doctors rate their colleagues more favourably when compared to feedback from nurses.¹⁴ Although this potential problem has been in part circumvented, at least in theory, in mini-PAT by producing guidelines to trainees on choosing raters, there remains a selection bias which may be reduced if the list of raters is discussed and agreed by both supervisor and trainee beforehand.

The process of peer review is a new 'culture', where individuals are asked to make a judgement on the performance of colleagues. Some may not feel comfortable with this process while others may not fully appreciate the importance of the objective nature involved in this assessment. Education about the process may enhance credibility of the tool²⁴ and reduce errors of measurement such as halo effect and central tendency.^{15, 25} Without adequate education, colleagues may not give a reliable opinion.

Feedback

With all multi-source assessment tools, considerable emphasis has been placed on the process and quality of feedback given by the educator.^{26, 27} Therefore the educational supervisor should be

appropriately trained for the task. The training should not be limited to interpreting numerical data in the feedback report, as this has been shown to be inadequate to identify improvement needs,²⁸ but it is vital that the educator is adequately trained to adopt an effective and constructive approach when discussing the results of the report. It has been shown that non-specific feedback does little to effect a change in performance.²⁸ The educational supervisor should also be trained to allow learner interaction, and be reasonably familiar in putting together, with the learner, an action plan for development. He/she should be told not to dwell on isolated negative incidences from the free text comments. The supervisor should be aware that comments, though interesting, often relate to recent incidents and may have a negative impact on motivation. In one study half the doctors who received negative feedback questioned its validity and did not accept or use it.²⁸

Without adequate training, therefore, feedback is entirely dependent on the skill of educator, personal interest and previous experience. One way of improving the feedback process is by offering training courses in feedback to educational supervisors. This is understandably a large task and requires time but in the interim it may be worthwhile to limit the responsibility of feedback in any one hospital to only a few educational supervisors who have undergone adequate training. The manner of giving feedback is crucial if the aim is improved conduct.

Relating mini-PAT to patient outcome

Ramsey noted that there is no evidence that peer rating improves or predicts patient outcome. However it is plausible that good communication skills improve patient satisfaction. The latter could be looked at separately through a different tool and it may be worthwhile linking the two (patient satisfaction and mini-PAT) to each other to provide firm evidence for validation that mini-PAT is a useful tool to determine improved patient care.

Conclusion

This assessment tool is still in its infancy and further research and development is needed. As experience is acquired, it is crucial to review the process and refine the tool being used. The end result should be better patient care and fair, equitable treatment for doctors through providing the necessary information they need for their professional development.

References

- 1 Southgate L, Cox T, David T, *et al*. The General Medical Council's performance procedures: peer review of performance in the workplace. *Med Educ* 2001;**35**(Suppl 1):9-19
- 2 Matthews BP, Redman T. The attitudes of service industry managers toward upward appraisal. *Career Develop Int* 1997;**2**:46-53
- 3 Atwater LA, Waldmand D, Atwater RD, Cartier A. An upward feedback field experiment. Supervisor's cynicism, follow-up and commitment to subordinates. *Personnel Psychology* 2000;**53**:275-97
- 4 Kluger AN, Denisi A. The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback theory. *Psychol Bull* 1996;**119**:254-84
- 5 Bret J, Atwater L. 360-degree feedback: accuracy, reactions and perceptions of usefulness. *J Appl Psychol* 2001;**86**:930-42
- 6 Blurton RR, Mazzaferri EL. Assessment of interpersonal skills and humanistic qualities in medical residents. *Med Educ* 1985;**60**:648-50
- 7 Linn LS, Dimatteo MR, Cope DW, Robbins A. Measuring physicians' humanistic attitudes, values, and behaviours. *Med Care* 1987;**25**:504-13
- 8 Keynan A, Freidman M, Benbassat J. Reliability of global rating scales in the assessment of clinical competence of medical students. *Med Educ* 1987;**21**:477-81
- 9 Carline JD, Paauw DS, Thiede KW, Ramsey PG. Factors affecting the reliability of ratings of students' clinical skills in a medicine clerkship. *J Gen Intern Med* 1992;**7**:506-10
- 10 Glasscock RJ, Benson JA, Copeland RB, *et al*. Time-limited certification and recertification: the program of the American Board of Internal Medicine. *Ann Intern Med* 1991;**114**:59-62
- 11 Ramsey PG, Wenrich MD, Carline JD, Inui TS, Larson EB, Logerfo JP. Use of peer ratings to evaluate physician performance. *JAMA* 1993;**269**:1655-60
- 12 Violato C, Lockyer J, Fidler H. Multisource feedback: a method of assessing surgical practice. *BMJ* 2003;**326**:546-8
- 13 Joshi R, Ling F, Jaeger J. Assessment of 360-degree instrument to evaluate residents competency in interpersonal and communication skills. *Acad Med* 2004;**79**:458-63
- 14 Johnson D, Cujec B. Comparison of self, nurse and physician assessment of residents rotating through an intensive care unit. *Crit Care Med* 1998;**26**:1811-6
- 15 Norcini JJ. Peer assessment of competence. *Med Educ* 2003;**37**:539-43
- 16 Ramsey PG, Carline JD, Blank LL, Wenrich MD. Feasibility of hospital-based use of peer ratings to evaluate the performance of practising physicians. *Acad Med* 1996;**71**:364-70
- 17 Hall W, Violato C, Lewkonja R, *et al*. Assessment of physician performance in Alberta the physician achievement review. *CMAJ* 1999;**161**:52-7
- 18 Archer JC, Norcini J, Davies HA. Use of SPRAT for peer review of paediatricians in training. *BMJ* 2005;**330**:1251-3
- 19 General Medical Council. *Good medical practice*. 2001 See <http://www.gmc-org.uk>
- 20 Davies H, Archer J. Multi source Feedback: development and practical aspects. *Clin Teacher* 2005;**2**:77-81
- 21 Southgate L, Grant J. *Principles of an assessment system for postgraduate medical training*. London: PMETB, 2004 A working paper from the Postgraduate Medical Education Training Board
- 22 The Foundation Programme Committee of the Academy of Medical Royal Colleges, in co-operation with Modernising Medical Careers in the Departments of Health. *Curriculum for the Foundation Years in Postgraduate Education and Training*. London: Department of Health, 2005
- 23 Lurie SJ, Nofziger AC, Meldrum S, Mooney C, Epstein RM. Effects of rater selection on peer assessment among medical students. *Med Educ* 2006;**40**:1088-97
- 24 Bracken DW, Timmreck CW, Church AH, editors. *The Handbook for Multisource Feedback: the comprehensive resource for designing and implementing MSF processes*. San Francisco, CA: Jossey-Bass, 2001
- 25 Johnson D, Cujec B. Comparison of self, nurse, and physician assessment of residents rotating through an intensive care unit. *Crit Care Med* 1998;**26**:1811-6
- 26 Bing-You RG, Paterson J. Feedback falling on deaf ears: residents' receptivity to feedback tempered by sender credibility. *Med Teacher* 1997;**19**:40-4
- 27 Sargeant J, Mann K, Ferrier S. Exploring family physicians' reactions to multisource feedback: perceptions of credibility and usefulness. *Med Educ* 2005;**39**:497-504
- 28 Sargeant J, Mann K, Sinclair D, van der Vleuten C, Metsemakers J. Challenges in multisource feedback: intended and unintended outcomes. *Med Educ* 2007;**41**:583-91