

**Ajit Narayanan**  
is Head of the School of Computing and Mathematical Sciences at Auckland University of Technology, New Zealand. His previous posts were Head of Computer Science and Dean of Science at the University of Exeter, and Head of Computing, University of Portsmouth (both in the UK).

He has worked as a statistical consultant on a revalidation feasibility study with the UK General Medical Council involving the Peninsula Medical School and CFEP. His research interests are in the application of machine learning techniques to cancer gene expression data.

**Michael Greco**  
is Director of CFEP-UK Surveys (University of Exeter) which is a lead organisation in administering the patient questionnaires in the new British GP Contract. His academic appointments are with the Peninsula Medical School (UK) and the School of Medicine, Queensland (Australia). His previous posts were with the National Primary Care Development Team (Director of Patient Experience) and the National NHS Clinical Governance Support Team (Communication Skills Development Lead).

**Keywords:** 360 appraisal, questionnaires, performance assessment, multi-source feedback, peer assessment, clinical ability, professional competence, cluster analysis

Professor Ajit Narayanan  
Head of School of Computing and Mathematical Sciences  
Auckland University of Technology, AUT Tower,  
2-14 Wakefield Street  
Auckland 1142,  
New Zealand  
Tel: +64 9 921 9345  
E-mail: ajit.narayanan@aut.ac.nz

# What distinguishes general practitioners from consultants, according to colleagues?

Ajit Narayanan and Michael Greco

Received (in revised form): 3rd April 2007

## Abstract

There is still very little understanding of what exactly is being measured when peer review is undertaken and how it should be measured for different target groups. The aim of this study is to determine the basis on which raters distinguish between two sets of colleagues: GPs and consultants. From March to July 2005 a study was run to test a newly created peer assessment tool for doctors in primary and secondary care settings. A total of 33 doctors took part — 14 GPs and 19 hospital consultants. Each of the participants was asked to identify 15 colleagues as appraisers to complete the colleague feedback evaluation tool questionnaire. Factor analysis demonstrates that colleagues have rated all doctors along the four dimensions of effective communication, clinical competence, time-management and trust. After separating GP colleague responses from consultant colleague responses, the application of cluster analysis clearly demonstrates that overall ability as a GP is strongly linked to effective communication with colleagues, whereas for consultants overall ability is based mainly on clinical competence alone. The implications of these findings for the development of future peer evaluation tools are discussed.

## Introduction

There is growing interest in the UK and internationally in the role of peer assessment for the purpose of appraisal, revalidation and clinical governance.<sup>1,2</sup> One major US instrument is the ‘professional associate rating’ (PAR) developed by Ramsey *et al.*<sup>3</sup> for hospital general internists, consisting of 11 items (nine-point Likert-type scale), four of which were respect, integrity, compassion and responsibility. Eighty-nine per cent of variation could be ascribed to two factors: cognitive/clinical (74.6 per cent) and humanistic (14.1 per cent). Another US instrument is the ‘peer review evaluation form’ (PREF) developed by Thomas *et al.*<sup>4</sup> for hospital internal residency training, consisting of a number of items (technical skills, interpersonal skills) using nine-point Likert-type scales. The two factors accounted for about 47–55 per cent and 29–41 per cent, respectively, of the variance. The Canadian ‘peer assessment questionnaire’ (PAQ) developed by Hall *et al.*<sup>5</sup> for mainly family

---

*What distinguishes general practitioners from consultants, according to colleagues?*

physicians and clinical specialists uses five-point Likert-type scales for 24 items. Seventy-five per cent of total variation could be ascribed to four factors: clinical competency, professional management, humanistic communication and psychological management. The Sheffield peer assessment tool (SPRAT)<sup>6</sup> is a recent UK addition designed to assess components of trainee paediatricians. The questionnaire consists of 24 questions covering good clinical care; maintaining good medical practice; teaching and training, assessing and appraising; relationships with patients; and working with colleagues. A six-point scale was used, ranging from 1 (very poor) to 6 (very good).

One of the aims of this paper is to demonstrate that the colleague feedback evaluation tool (CFET) provides an additional useful instrument for peer review across a range of doctors in primary and secondary care settings. Another aim is to determine whether CFET can be used to obtain information and knowledge of concepts underlying peer evaluation of doctors and consultants to guide the design of future peer evaluation tools. The final aim is to demonstrate how cluster analysis reveals conceptual groupings among the questionnaire items, thereby making explicit how raters make judgments on colleagues.

## **Methods**

A draft instrument was developed based on existing literature (including the General Medical Council's *Good Medical Practice* guidelines)<sup>7</sup> followed by the Delphi Technique (a structured approach for establishing a consensus opinion from experts) with 11 doctors. This process resulted in a questionnaire consisting of 18 five-point Likert scale items, covering the following: clinical knowledge, clinical ability, communication with patients, compassion/empathy, colleague communication, punctuality and reliability, respect for colleagues, ability to say 'no', awareness of limitations, team orientation, use of resources, ability to manage stress, confidentiality, appearance and behaviour, respect to their own health, trustworthiness/honesty/probity, management/leadership skills, and overall ability as a doctor. A total of 33 doctors took part in the subsequent survey — 13 GPs and 21 hospital consultants, whose specialisms included orthopaedics, dermatology, orthodontics, ear-nose-throat, oral and maxillofacial surgery, anaesthetics, upper gastrointestinal surgery, general surgery, infectious diseases and respiratory medicine. Each of the participants was asked to identify 15 colleagues as appraisers.

With regard to statistical methods, reliability using Cronbach's  $\alpha$  was adopted. Factor analysis (principal component analysis) was used to investigate whether there was any reduction possible in the number of questionnaire items (data reduction) as well as to detect structure in the relationships between items (structure detection). 'Cluster analysis'<sup>8</sup> is a generic term that covers a number of methods for finding groupings in the data, where the number of groups and their component parts are not known in advance. It is used to explore and mine the data with the purpose of categorising different items into groups such that the degree

of association between two items is maximal (their variance is minimal) if they belong to the same group and minimal (their variance is maximal) otherwise. The output is typically in the form of a taxonomy or hierarchical tree (dendogram). A number of different distance measures are in use to identify similarity/dissimilarity of items and Pearson correlation was used as the similarity measure in this study.

## Results

### First-level analysis

An average of 13.46 appraisers returned questionnaires for each GP and 12.52 for each consultant. The mean of all mean values of the 18 items (316 colleague questionnaires for 13 GPs and 21 consultants) was 4.164, with minimum 3.673 (V12: ability to manage stress) and maximum 4.548 (V14: appearance and behaviour). Reliability analysis on all 18 questionnaire items resulted in high scores (Cronbach  $\alpha = 0.93$ ).

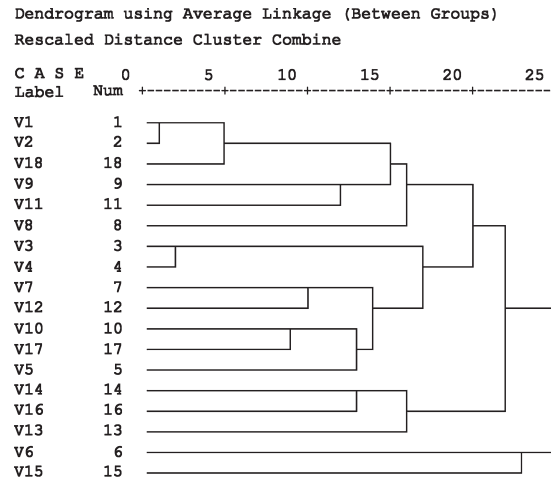
Inter-item correlations between each of the 18 items averaged 0.426 (minimum 0.154, maximum 0.79, all statistically significant  $\leq 0.01$ ), and the average item-total correlation averaged 0.636. Strong inter-item correlations were confirmed by clustering all 18 items (V1–V18).

Items 1 and 2 (joined at distance 1.5, top left part of Figure 1) relate to clinical knowledge and clinical ability, and item 18 (joining items 1 and 2 at distance 5) deals with overall ability as a GP/consultant. These three items are the most strongly related to each other (as determined by the rescaled distance metric at the top of Figure 1). Items 3 and 4 (joined at distance 2 and in a separate cluster) deal with compassion/empathy and communication with colleagues. Items 10 and 17 (joined at distance 8.5) deal with team orientation and management/leadership skills, while items 7 and 12 (joined at distance 9) deal with respect for colleagues and ability to manage stress.

Towards the top levels of the tree (right part of Figure 1), it appears that the scale of 18 items can be split into two basic parts (distance 25): items 6 and 15 on the one hand (punctuality/reliability and respect for one's health) and the remaining items.

Factor analysis (principal component analysis using varimax) was run on the total data set and four factors were found (eigenvalues  $\geq 1$ ). The first factor loads most heavily on V4, V7 and V10; the second on V1, V2 and V18; the third on V8 and V6; and the fourth on V13 and V14. There provides some evidence that compassion/empathy (V4), respect for colleagues (V7) and team orientation (V10) form one substructure ('effective communication'); that clinical knowledge (V1), clinical ability (V2) and overall ability as a GP/consultant (V18) form a second substructure ('clinical competence'); that ability to say 'no' (V8) and punctuality/reliability (V6) form a third substructure ('time management'); and that respect for confidentiality (V13) and appearance/behaviour (V14) form another substructure ('professionalism'). The second factor (V1, V2 and V18) clearly reflects the most strongly associated subcluster of items found through cluster analysis.

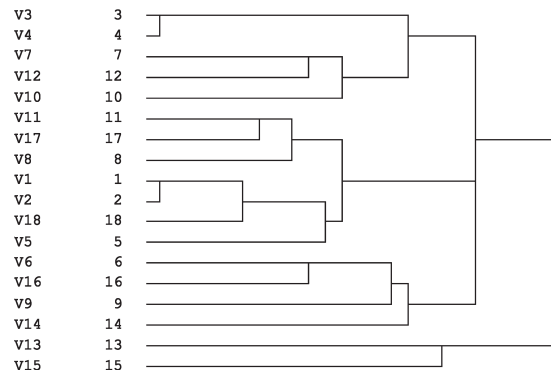
## What distinguishes general practitioners from consultants, according to colleagues?



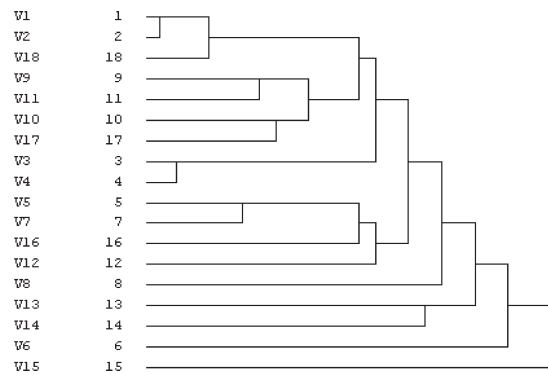
**Figure 1:** Cluster analysis of all 18 items for all colleague responses  
Dendrograms (hierarchical tree diagrams) display the relative size of the correlation coefficients at which variables are combined. The stronger the coefficient between two variables, the less distance there is between them and the more likely it is that they will be combined towards the left of the diagram above. The dendrogram above is depicted horizontally, with each row (CASE label) representing a variable. The line 0–25 at the top of the diagram is a ‘dissimilarity’ measure and represents a rescaled version of the correlation coefficients, with strongest similarity at 0 and least similarity at 25. Variables with strong similarity are close together towards the left of the diagram. The lines represent the agglomeration order in which individual variables combine into pairs, triples, quadruples, etc., until all variables are combined (at the right-hand side of the diagram). The ‘average linkage (between groups)’ title at the top of the diagram indicates that, as variables are combined using correlation coefficients, a new coefficient is calculated for the agglomerated combination that is the average of the coefficients between the variables making up that agglomeration. This new coefficient is then used to decide which other variable or combination to agglomerate next. See the main text for interpretations of the dendrogram. SPSS V12 was used for all statistical analysis.

The data were then split between GPs and consultants. The reliability scores showed a small difference between GPs and consultants (0.906 for GPs as opposed to 0.935 for consultants), but this difference is within an acceptable 5 per cent interval range (ie 95 per cent confidence interval) for the GPs and consultants as a whole.

Cluster analysis of the way that questionnaire items agglomerated for the two different populations of GPs (Figure 2) and consultants (Figure 3) demonstrated some basic similarities but also showed some interesting differences.



**Figure 2:** Cluster analysis of all 18 rating items for GP colleague responses only, showing strong relationships between V3 and V4 as well as between V1 and V2



**Figure 3:** Cluster analysis of all 18 rating items for consultant colleague responses only, showing strong relationships between V1, V2 and V18 as well as between V3 and V4

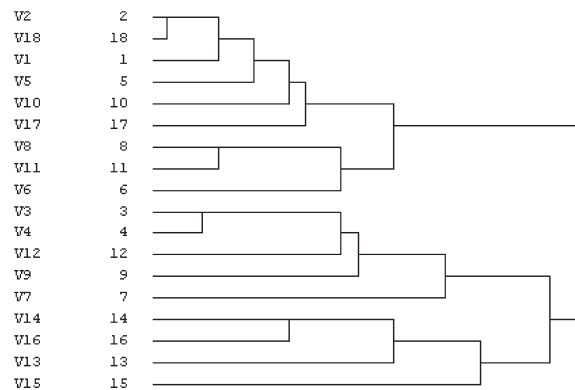
Colleagues' perceptions of GPs (Figure 2) demonstrate equal importance (as given by the distance metric) between communication and compassion with patients in one cluster (V3 and V4, respectively) on the one hand, and clinical knowledge (V1) and clinical ability (V2) on the other in another cluster. Colleagues' perceptions of the individual GP's overall ability as a doctor (V18) were most strongly associated with V1 and V2 and also with communication with colleagues (V5). Additionally, compassion and communication with patients are more strongly related to respect for colleagues (V7), ability to manage stress (V12) and team orientation (V10) than to clinical ability.

For consultants, colleagues' priorities appear to be different (Figure 3). The most important cluster for consultants is clinical knowledge and clinical ability (V1 and V2) in one cluster, followed by communication and compassion (V3 and V4) in another. Next in terms of importance appeared to be communication with and respect for colleagues (V5 and V7). Consultant colleagues associated overall ability as a consultant most strongly with clinical knowledge and clinical ability by themselves (rather than with communication with colleagues, as is the case with GP colleagues). Additionally, for GP colleagues, V7 (respect for colleagues) was clustered with V12 and V10 (ability to manage stress and team orientation, respectively), rather than with V5 (communication with colleagues), as is the case for consultant colleagues.

### Second-level analysis

The data were then aggregated by GP and consultant (ie all colleague responses for the 18 questionnaire items for each GP/consultant were summed and averaged, resulting in 18 averaged scores for that GP/consultant) for a second-level analysis of the data. For the 14 GPs, 175 colleague questionnaires were aggregated by each GP. The reliability for these aggregated items was a high 0.947. Cluster analysis of the aggregated GP data (Figure 4) showed major similarities with the unaggregated cluster analysis for GP colleague responses (Figure 2), thereby demonstrating that there is no difference between GP responses at a population or individual level. Again, GP colleagues related overall

*What distinguishes general practitioners from consultants, according to colleagues?*



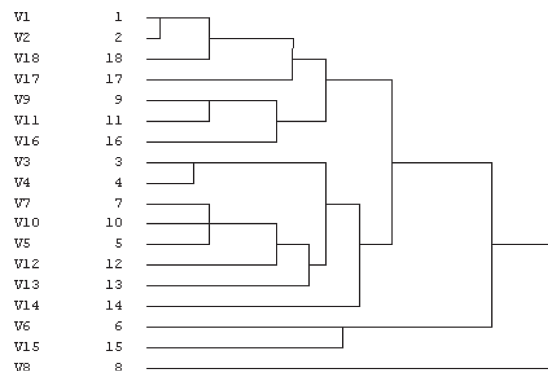
**Figure 4:** Cluster analysis of all 18 rating items for aggregated GP colleague responses only, showing strong relationships between V2, V18 and V1, between V3 and V4, and between V8 and V11

ability as a doctor (V18) most strongly with clinical knowledge and clinical ability (V1 and V2), and the relationship between communication with patients and compassion still came through.

For the 19 hospital consultants, 263 consultant colleague questionnaires were aggregated for each consultant. The reliability coefficient was high at 0.944. Cluster analysis of the aggregated consultant data again revealed major similarities with the unaggregated results (Figure 5), with the strong relationships between V1, V2 (clinical knowledge and ability) and V18 (overall ability) still clearly visible.

## Discussion

The results of the cluster analysis at both the unaggregated and aggregated levels confirm that CFET is robust at both levels of analysis in terms of eliciting responses from colleagues that can be trusted to represent a repeatable measurement of (a) GPs as individuals as well as GPs as a group (Figures 2 and 4), and (b) hospital consultants as individuals as well as hospital consultants as a group (Figures 3 and 5).



**Figure 5:** Cluster analysis of all 18 rating items for aggregated consultant colleague responses only, showing strong relationships between V1 and V2, between V3 and V4, between V9 and V11 and between V7, V10 and V5

The results support the finer-grained distinctions in underlying conceptual structures as reported by PAQ, which also found four factors: clinical competency, professional management, humanistic communication and psychological management. Of the four factors found by CFET, clinical competence is shared. However, the other three factors found by CFET are not only refinements of professional management, humanistic communication and psychological management but also extensions of our understanding of what makes a good doctor, according to colleagues.

The cluster analysis when GP responses are separated from hospital consultant responses reveals subtle differences in the way that GPs are perceived and rated by hospital consultants and how the various attributes are linked together. In particular, by separating GP colleague responses from consultant colleague responses, the analysis has demonstrated that overall ability as a GP is most strongly linked to effective communication with colleagues, whereas for consultants, overall ability appears to be mainly based on clinical competence alone.

## Conclusion

One interesting aspect of the dendograms for GPs and consultants at both the unaggregated and aggregated levels is that there is a clear separation between overall ability as a doctor (as associated with clinical knowledge and ability) and communication skills with patients. Communication skills are increasingly being seen as essential for any doctor speciality, as they have an impact on health outcomes (eg functional and emotional status<sup>9,10</sup>), concordance with treatment and improved trust.<sup>11,12</sup> A number of medical professional bodies, including the General Medical Council, Academy of Royal Medical Colleges, Royal College of General Practitioners and Royal College of Physicians, have recently highlighted the importance of doctors developing good communication skills.<sup>13</sup> Benefits that can result when doctors practise good and effective communication skills with their patients include patients' problems being more easily identified and fewer clinical errors being made.<sup>14</sup> The results presented here demonstrate that there is awareness concerning the importance of good and effective communication as an essential skill. However, it appears from the cluster analysis that colleagues do not yet relate good patient communication skills with the clinical competence of their doctor colleagues.

## References

1. Norcini, J. (2003) 'Peer assessment of competence', *Medical Education*, Vol. 37, No. 6, pp. 539–543.
2. Evans, R., Elwyn, G. and Edwards, A. (2004) 'Review of instruments for peer assessment of professionals', *British Medical Journal*, Vol. 328, No. 7450, pp. 1240–1245.
3. Ramsey, P. G., Carline, J. D., Blank, L. L. and Wenrich, M. D. (1996) 'Feasibility of hospital-based use of peer ratings to evaluate the performance of practicing physicians', *Academic Medicine*, Vol. 71, No. 4, pp. 364–370.
4. Thomas, P., Gebo, K. and Hellmann, D. (1999) 'A pilot study of peer review in residency training', *Journal of General Internal Medicine*, Vol. 14, No. 9, pp. 551–514.

---

*What distinguishes general practitioners from consultants, according to colleagues?*

5. Hall, W., Violato, C., Lewkonina, R., Lockyer, J., Fidler, H., Toews, J., Jennett, P., Donoff, M. and Moores, D. (1999) 'Assessment of physician performance in Alberta: The Physician Achievement Review', *Canadian Medical Association Journal*, Vol. 161, No. 1, pp. 52–57.
6. Archer, J. C., Norcini, J., Davies, H. A. (2005) 'Use of SPRAT for peer review of paediatricians in training', *British Medical Journal*, Vol. 330, No. 7502, pp. 1251–1253.
7. General Medical Council (2006) *Good Medical Practice*, General Medical Council, London.
8. Kauffman, L. and Rousseeuw, P. J. (1990) *Finding Groups in Data: An Introduction to Cluster Analysis*, John Wiley and Sons, Inc, New York.
9. Maly, R. L., Bourque, L. and Engelhardt, R. F. (1999) 'A randomized control trial of facilitating information giving to patients with chronic medical conditions: Effects on outcomes of care', *Journal of Family Practice*, Vol. 45, No. 5, pp. 356–363.
10. Roter, D. L., Hall, J. A., Kern, D. E., Barker, L. R., Cole, K. A. and Roca, R. P. (1995) 'Improving physicians' interviewing skills and reducing patients' emotional distress. A randomized clinical trial', *Archives of Internal Medicine*, Vol. 155, No. 17, pp. 1877–1884.
11. Stewart, M., Brown, J. B., Donner, A., McWhinney, I. R., Oates, J., Weston, W. W. and Jordan, J. (2000) 'The impact of patient-centered care on outcomes', *Journal of Family Practice*, Vol. 49, No. 9, pp. 796–804.
12. Hausman, A. (2001) 'Taking your medicine: Relational steps to improving patient compliance', *Health Marketing Quarterly*, Vol. 19, No. 2, pp. 49–71.
13. General Medical Council (2001) *Good Medical Practice*, General Medical Council, London.
14. Maguire, P. and Pitceathly, C. (2002) 'Key communication skills and how to acquire them', *British Medical Journal*, Vol. 325, No. 7366, pp. 697–700.