8 Saturno P, Heather Palmer R. Physician attitudes, self-estimated performance and actual compliance with locally peer-defined quality evaluation criteria. *Int J Qual Healthcare* 1999;11:487-96.
9 Ramsey P, Carline JD, Blank LL, Wenrich MD. Feasibility of hospital-based use of peer ratings to evaluate the performance of practicing physicians. *Acad Med* 1996;71:364-70.
10 Lipner R, Blank L. The value of patient and peer ratings in recertification. *Acad Med* 2002;77:S64-6.
11 Hall W, Violato C, Lewkonia R, Lockyer J, Fidler H, Toews J, et al. Assessment of physician performance in Alberta: the physician achievement review. *CMAJ* 1999;161:52-6.
12 Thomas P, Gebo K, Hellmann D. A pilot study of peer review in residency training. *J Gen Intern Med* 1999;14:551-4.
13 American Board of Internal Medicine. *Recertification: program for continuous professional development.* www.abim.org/cpd/cpdhome/index.htm (accessed 23 Apr 2004).
14 Streiner DL, Norman GR. *Health measurement scales: a practical guide to their development and use.* Oxford: Oxford University Press, 1995.
15 Ramsey P, Carline JD, Inui TS, Larson EB, LoGerfo JP, Wenrich MD. Predictive validity of certification by the American Board of Internal Medicine. *Ann Intern Med* 1989;110:719-26.
16 Fidler H, Lockyer JM, Toews J, Violato C. Changing physicians' practices: the effect of individual feedback. *Acad Med* 1999;74:702-14.
17 Peirperl M. Conditions for the success of peer evaluation. *Int J Hum Resour Manage* 1999;10:429-58.
18 Atwater L. Self-other agreement: does it really matter? *Personnel Psychology* 1998;51:577-98.
19 Festinger L. A theory of social comparison processes. *Hum Relat* 1954;7:117-40.
20 Fox S, Ben-Nahum Z. Perceived similarity and accuracy of peer ratings. *J Applied Psychol* 1989;74:781-6.
21 Mumford M. Social comparison theory and the evaluation of peer evaluation: a review and some applied implications. *Personnel Psychology* 1983;36:867-81.
22 Church A. Do you see what I see? An exploration of congruence in ratings from multiple perspectives. *J Applied Soc Psychol* 1997;27:983-1020.
23 Kaplan C, Centor R. The use of nurses to evaluate houseofficers' humanistic behaviour. *J Gen Intern Med* 1990;5:410-4.
24 Davis J. Comparison of faculty, peer, self, and nurse assessment of obstetrics and gynecology residents. *Obstet Gynecol* 2002;99:647-51.
25 Van den Bos K. Procedural and distributive justice: what is fair depends more on what comes first than on what comes next. *J Pers Soc Psychol* 1997;72:95-104.
26 Vuorinen R, Tarkka M. Peer evaluation in nurses' professional development: a pilot study to investigate the issues. *J Clin Nurs* 2000;9:273-81.

(Accepted 30 January 2004)

Commentary: "Soft" assessment—an oxymoron?

Murray Lough

NHS Education for Scotland, Glasgow G3 8BW
Murray Lough
assistant director
murray.lough@nes.scot.nhs.uk

O wad some Power the giftie gie us,
To see oursels as ithers see us!

R Burns, 1798

The desire to assess the "softer" parts of practice is strong. In addressing this ambition, many of the tensions and controversies surrounding the assessment process are thrown into sharp focus. Exactly how one defines soft practice for the jobs carried out by, say, general practitioners or orthopaedic surgeons is open to wild speculation. One key message of the paper by Evans et al is the importance of those being assessed having a stake in their assessment; this may be the "missing link" that could make the assessment process less threatening.

Assessment has to be rigorous, not least because it should be fair. The subtleties of language highlight the difficulties—peer assessment and peer review have different connotations. In an era of individual accountability and clinical governance, many might question whether soft practice can ever be assessed with rigour.

Evans and colleagues in their paper concentrate on what could currently be considered the psychometric gold standard attributes of assessment.[1] Van der Vleuten would argue that acceptability and cost effectiveness are also crucial in laying down a system where busy people are assessing busy people.[2] No matter how we cut the cake, assessment is expensive in terms of both time and money.

It is interesting that the three papers identified by Evans and colleagues emanate from North America, where a culture of a high level of psychometric skills is possibly encouraged by a litigation conscious community. This may highlight the lack of such skills in other countries and could serve as a warning that, given the high stakes involved in such assessments, the result is either instruments of poor quality, as suggested by the paper, or no instruments at all.

In response to this it may be that in considering assessment of a softer side of practice we need to consider some other measures of validity.[3] One example is consequential validity—that is, taking into consideration the consequences on those being assessed as part of the process.[4] In a world dominated by numbers and the desire to measure them, a wider debate about the strengths and limitations of formal psychometrics when considering the less tangible aspects of practice is urgently needed if a more valid assessment of doctors is to be achieved.

In considering how one might address the assessment of peers, particularly with reference to the complexities of the jobs done, one example is to consider assessments over time—that is, trends such as the longitudinal evaluation of performance (LEP) used for the assessment of dental trainees in Scotland.[5] It covers a range of competencies on a nine point scale and concentrates on feedback, a crucial element in consequential validity.

Three hundred and sixty degree or multisource feedback provides judgments on strengths and areas for potential improvement as perceived by work colleagues both above and below in the hierarchy. Many see this very practical method as opening new opportunities in exploring humanistic qualities that are not easily assessed by more traditional methods.[6] Predictably, however, there are limitations, not least with validity, but these are not insurmountable.

Perhaps more qualitative work on defining soft practice is required before designing any more instruments. Themes on areas of soft practice could be collated drawn from groups of doctors with additional input from patients to form the basis for more formal psychometric development and testing. Without this, soft assessment will be seen as a soft option and not given the place it deserves.

Competing interest: None declared.

1 Evans R, Elwyn G, Edwards A. Review of instruments for peer assessment of physicians. *BMJ* 2004;328:1240-3.
2 Van der Vleuten CPM. The assessment of professional competence: developments, research and practical implications. *Adv Health Sci Educ* 1996;1:41-67.
3 Messick S. The psychology of educational measurement. *J Educ Measur* 1984;21:215-37.
4 Moss PA. Shifting conceptions of validity in educational measurement: implications for performance assessment. *Rev Educ Res* 1992;62:229-58.
5 Prescott LE, Norcini JJ, McKinlay P, Rennie JS. Facing the challenges of competency-based assessment of postgraduate dental training, longitudinal evaluation of performance (LEP). *Med Educ* 2002;36:92-7.
6 King J. 360° appraisal. *BMJ* 2002;324:S195-6.