

## Patient feedback in revalidation:

an exploratory study using the consultation satisfaction questionnaire

### Abstract

#### Background

Revalidation is the UK process for the review of doctors to ensure they are fit to practise. Revalidation will include patient feedback.

#### Aim

To investigate the role of patient feedback on GPs' consultations in revalidation.

#### Design and setting

Cross-sectional survey of patients consulting 171 GPs.

#### Method

A total of 6433 patients aged 16 years or over completed the consultation satisfaction questionnaire (CSQ). Generalisability analysis was undertaken, scale scores calculated, and outliers identified using two and three standard deviations from the mean as control limits. Comments made by patients were categorised into positive, neutral, or negative.

#### Results

After averaging each scale for each doctor, mean scores (standard deviation), out of a possible score of 100, were: general satisfaction 78.1 (7.2); professional care 82.1 (6.1); relationship 71.2 (7.1); perceived time 65.7 (7.6). A D-study (which enables estimation of the reliability from 0–1 of the CSQ scores for different numbers of responders for each doctor), indicated that ratings by 19 patients would achieve a generalisability coefficient of 0.80 for the combined score. Fifteen GPs had one or more scale scores below two standard deviations of the mean. Comments were more often negative for GPs with scores below two standard deviations of the mean.

#### Conclusion

Most patients of most GPs are satisfied with their experience of consultations, and ways to make patient feedback formative for these doctors is required. For a few GPs, most patients report some dissatisfaction. Patient feedback may identify doctors who need educational support and possibly remediation, but agreed questionnaire score thresholds are required, and agreement is needed on the weight to be attached to patient experience in comparison with other aspects of performance.

#### Keywords

certification; general practice; patient satisfaction.

### INTRODUCTION

Revalidation is the process by which doctors in the UK will demonstrate to the regulatory authority, the General Medical Council (GMC), that they remain up to date and fit to practise.<sup>1</sup> Revalidation will not begin before completion of extensive pilot tests. It is intended that it should rely in large part on systematic annual appraisal of doctors, including review of information about the doctor's performance in relation to standards set out by the GMC, which will be assessed by methods that include feedback from colleagues and patients.<sup>2</sup>

The Royal College of General Practitioners (RCGP) has responsibility for proposing the standards and methods of revalidation for GPs, subject to approval by the GMC.<sup>3</sup> The RCGP currently advises that once revalidation is introduced, the portfolio of evidence submitted by GPs when seeking revalidation should include the results of a patient survey.<sup>4</sup> The use of surveys is expected to identify a small number of doctors with low ratings on patient experience of, or satisfaction with, care, and a process of investigation and, if necessary, remediation is planned for these doctors.<sup>5</sup> Outstanding issues to consider include whether poor patient experience scores alone should be considered sufficient evidence to trigger such investigation, or whether evidence of poor performance across a range of measures should be required; the threshold scores to be adopted in deciding on the need for investigation,

educational support, or even remediation; and the numbers of doctors likely to be found in need of investigation. Therefore, a study of the potential role of surveys of patient experience in revalidation was undertaken to stimulate debate on the use of patient experience in revalidation.

### METHOD

This study used data already collected by GPs for use in appraisal, using the consultation satisfaction questionnaire (CSQ). The GMC has published criteria for the approval of questionnaires,<sup>4</sup> and has devised and piloted colleague and patient questionnaires.<sup>6</sup> The RCGP has published a report comparing instruments that might be used.<sup>7</sup> The CSQ is an 18-item instrument designed to provide information on patient satisfaction with consultations with GPs, and is being assessed for use in revalidation, although it is not yet approved.<sup>8</sup> It has been extensively evaluated, being shown to possess good reliability and validity,<sup>8,9</sup> and has been used in independent research studies.<sup>10,11</sup>

The CSQ includes questions about the consultation grouped into four scales: three questions dealing with general satisfaction with the consultation, seven with experience of professional aspects of care (the examination, information about treatment, being treated as a person), five with the relationship (being able to disclose personal information, the doctor understood the patient), and three with the length of the

**R Baker**, MD, FRCGP, professor; **C Tarrant**, BSc, PhD, lecturer in health psychology; **N Taub**, PhD research fellow in Medical Statistics, Department of Health Sciences, University of Leicester, Leicester. **A Smith**, MSc, survey director, PatientDynamics, High Wycombe; **R McKinley**, MD, FRCP, FRCGP, professor of academic general practice, Keele University School of Medicine, Keele.

#### Address for correspondence

Richard Baker, Department of Health Sciences,

University of Leicester, 22–28 Princess Road West, Leicester, LE1 6TP.

**E-mail:** rb14@le.ac.uk

**Submitted:** 7 February 2011; **Editor's response:** 21 March 2011; **final acceptance:** 10 May 2011.

©British Journal of General Practice

This is the full-length article (published online 26 Sep 2011) of an abridged version published in print. Cite this article as: **Br J Gen Pract 2011; DOI: 10.3399/bjgp11X601343.**

### How this fits in

Revalidation for GPs in the UK will include feedback from patients, but there is a lack of information on how feedback should be used to identify GPs whose performance would benefit from investigation, support, or even remediation. Using the findings from patient surveys for 171 GPs, this study identified a small number of GPs with low scores on patient feedback. Systems are needed to ensure that all GPs are able to reflect on their feedback, and to differentiate between those GPs who might need formal remediation and those who might benefit from consultation skills training in the context of continuing education and appraisal.

consultation (sufficient time to deal with everything the patient wanted). The response to each question is in a five-point 'strongly agree' to 'strongly disagree' format. Following recoding of negatively worded questions, a score out of a possible total of 100 is calculated for each scale, with a higher score indicating more positive ratings. A score below 50 would indicate that the majority of the responding patients of the doctor concerned are expressing a degree of dissatisfaction. The questionnaire also includes an open question, worded 'Do you have any other comments about the consultation?'

Since 2009, surveys using CSQ have been offered to GPs by PatientDynamics™, for use in appraisal.<sup>12</sup> The survey includes, for each participating GP, 50 copies of the questionnaire, instructions on administration of the survey, and independent analysis with feedback. The present study used the data from surveys completed by March 2010. The data were anonymised and therefore there is no descriptive information about the participating GPs or their practices, although patients were asked to record their age and sex on the questionnaire. GPs were instructed to invite adults aged 16 years or over who consecutively consulted them to complete a survey questionnaire. Patients were asked to complete the questionnaire following the consultation, with reference to the doctor they had just consulted. Since the study involved analysis of anonymous data collected as part of service evaluation, the local NHS ethics committee judged that NHS research ethics approval was not required. The researchers therefore sought and obtained ethics approval from the University of Leicester ethics committee.

Descriptive statistical analyses were undertaken in SPSS (version 16). In preliminary analyses, the level of non-response and distribution of responses to individual questions was investigated, and the internal consistency of the scales was checked. A generalisability (G)-analysis and D-studies [The D analysis enables estimation of the reliability of CSQ doctors' scores with different numbers of questionnaire responses for each doctor. G studies allow estimation of the reliability of judgements of a quality [in this case, satisfaction] made by multiple assessors [in this case, patients], of multiple subjects [in this case, doctors], across a number of encounters [in this case, consultations]].<sup>13</sup> were undertaken using G-String.<sup>14</sup> After checking the distribution of scores of the included doctors, upper and lower control limits of two and three standard deviations from the mean were set to highlight potential outliers.<sup>15,16</sup> The open comments were classified by one researcher according to whether they expressed positive, neutral, or negative views about the consultation, and these comments were related to the scores of doctors within or outside the control limits. Comments made about the practice rather than the consultation were excluded, although all comments made about the GP were included. To illustrate comments made about doctors with different scores, comments were randomly selected for inclusion in a table, by allocating a number to each comment and generating random numbers, stratified by whether or not scale scores were below two or three standard deviations of the mean.

### RESULTS

A total of 6433 questionnaires were completed, relating to a total of 171 UK GPs. There was a mean of 37.6 questionnaires per GP (range 19–50, a mean response rate of 75.2%). Of those patients who gave their sex, 2151 (35.0%) were men, and 3991 (65.0%) were women. The mean age of males was 52.9 years (standard deviation [SD] 18.7 years), females 48.7 years (SD 18.9 years).

The mean rate of non-response to individual questions was 4.4%, range 1.0–7.4%. Scale scores were not calculated for responders who had not answered one or more questions in a scale. The mean scores attained on the four scales of the CSQ are given in Table 1. The means differed across the scales, with the perceived time scale attracting the lowest mean score, and the professional care scale the highest. For each scale, the range between the highest

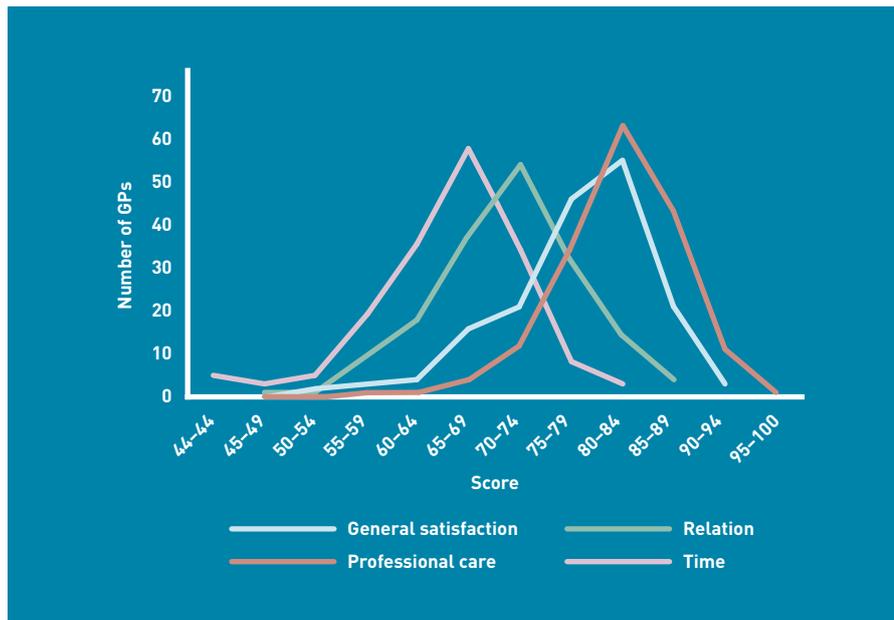
**Table 1. Scores for CSQ scales, showing mean, standard deviation, and minimum and maximum scores (n = 171 GPs)**

	General satisfaction	Professional care	Depth of relationship	Perceived time	Combined
Mean	78.1	82.1	71.2	65.7	75.8
Standard deviation	7.2	6.1	7.1	7.6	6.2
Minimum	52.5	58.5	48.3	40.2	53.8
Maximum	90.7	95.4	88.3	82.5	88.8

0.89, very similar to the findings for the questionnaire when first reported.<sup>8</sup> The D-analysis showed that, for the questionnaire's combined score (the average of the four scales, weighted by the number of the original 18 questions that they each use), reliability of 0.8 was achieved with 19 patients (Table 2). The G-coefficient was 0.84 for a harmonic mean of 24.7 ratings per doctor.

Table 3 gives the cut-off scores for upper and lower limits, using two and three standard deviations from the mean, and shows the numbers of GPs whose scores fell above the upper limits, and below the lower limits. For each scale, between six and eight doctors (3.5–4.7%) scored below the two standard deviation limit, and between one and four (0.6–2.3%) scored below the three standard deviation limit. Eleven (6.4%) scored between two and three standard deviations below the mean on at least one of the four scales. Of these 11, four (2.3%) scored below the two standard deviation limit on two of the scales, two (1.2%) scored below the limit on three of the scales, and one (0.6%) scored below the limit on all four scales. A total of four doctors scored below the three standard deviation limit for one or more scales, of whom one doctor scored below the limit for all four scales, and another scored below the limit for two scales. Fewer doctors scored above the upper control limits, with no doctor scoring above two standard deviations above the mean for general satisfaction, and no doctor scoring above three standard deviations for any of the scales.

A total of 1117 comments were made by patients in response to the open question (17.4% of responding patients). Of these



**Figure 1. Distribution of CSQ scale scores for 171 GPs.**

and lowest scores was around 40 scale points. The distributions of scores for all scales were close to normal, with a small tail evident for general satisfaction and professional care (Figure 1). Alpha coefficients for each scale were: general satisfaction 0.64, depth of relationship 0.76, perceived time 0.82, and professional care

**Table 2. Generalisability (G)-study and D-study findings, showing the effect of varying numbers of responses on the reliability of scores for subscales and the combined questionnaire**

	Mean number of patients per doctor	Generalisability coefficient				
		General satisfaction	Professional care	Depth of relationship	Perceived time	Combined <sup>a</sup>
G-analysis	24.7	0.800	0.812	0.812	0.722	0.841
D-analysis	13	0.678	0.695	0.695	0.578	0.736
	15	0.708	0.724	0.724	0.612	0.763
	17	0.734	0.748	0.749	0.641	0.785
	19	0.755	0.769	0.769	0.666	0.803
	21	0.773	0.786	0.786	0.688	0.818
	23	0.788	0.801	0.801	0.707	0.831
	25	0.802	0.814	0.814	0.724	0.843
	30	0.829	0.840	0.840	0.759	0.865

<sup>a</sup>The combined score was calculated as the mean of the four CSQ scales, weighted by the number of the original 18 questions that they each use.

**Table 1. Numbers of doctors out of 171 above or below two standard deviations (95%) and three standard deviations (99%), for scores of CSQ scales**

	General satisfaction		Professional care		Depth of relationship		Perceived time	
	Control limit score	Number of GPs	Control limit score	Number of GPs	Control limit score	Number of GPs	Control limit score	Number of GPs
<b>Above mean</b>								
+3 SDs	99.7	0	100	0	92.5	0	88.5	0
+2 SDs	92.5	0	94.3	1	85.4	3	80.9	3
<b>Below mean</b>								
-2 SDs	63.7	7	69.9	6	57.0	6	50.5	8
-3 SDs	56.5	2	63.8	2	49.9	1	42.9	4

**Table 4. Illustrative comments made by patients about consultations with doctors with no, or one or more, scale scores more than two or three standard deviations from the mean; comments selected randomly**

Above 2 standard deviations in all scales (n = 156)	Below 2 standard deviations in at least one scale (n = 11)	Below 3 standard deviations in at least one scale (n = 4)
<i>'One of the best doctors I have ever had.'</i>	<i>'I am very happy with the service I get.'</i>	<i>'The doctor was ignorant and rude.'</i>
<i>'So nice to be called in 10 minutes early.'</i>	<i>'Very good.'</i>	<i>'This doctor did not have time.'</i>
<i>'Made us very relaxed regarding very stressful and personal issues. Thank you.'</i>	<i>'Takes too long to sort my illness, has been over a year, not good enough.'</i>	<i>'Disgusted with the way I was treated.'</i>
<i>'Never had any trouble with my visits.'</i>	<i>'I was pleased with the consultation.'</i>	<i>'A doctor that is well liked and trusted by my family.'</i>
<i>'Lovely doctor.'</i>	<i>'Very good at her job. Excellent service.'</i>	<i>'Yes after visiting the doctor I booked another appointment to see a doctor that actually might find out what is wrong with me.'</i>

comments, 110 mentioned an aspect of the practice rather than the consultation and were excluded. Of the remaining 1007, most comments (975, 96.8%) were classified as positive or neutral. The 156 doctors with no scores below the two standard deviation limit received 948 comments, of which only 18 (1.9%) were negative; the 11 GPs with scores between two and three standard deviations of the mean received 34 comments, of which seven (20.6%) were negative; and the four doctors with scores below three standard deviations of the mean received 25 comments, of which seven (28.0%) were negative (difference between proportions  $P < 0.001$ ). Table 4 presents randomly selected example comments made about doctors with none, or one or more scale scores more than two, and more than three standard deviations from the mean.

## DISCUSSION

### Summary

This study found variation in scale scores

between doctors, but the majority of patients of the majority of doctors expressed satisfaction with their experience, as demonstrated by the scale scores and by the comments volunteered by responding patients. A small number of doctors achieved scores between two and three standard deviations below the mean, and some critical comments were made about them. A smaller number of doctors achieved scores below three standard deviations of the mean, and more than a quarter of the comments made about them were critical. The CSQ achieved adequate item response rates and demonstrated good internal consistency and satisfactory reliability. The G-coefficients were comparable to those of other instruments.<sup>6</sup>

### Strengths and limitations

This study used a questionnaire of established validity and reliability. Although a relatively large number of GPs were included, those that took part were

volunteers and are unlikely to be representative of all GPs in the UK. Furthermore, there is no information on their characteristics or their practices. Therefore, it is not possible to estimate the numbers of GPs who are likely to receive low scores in such surveys. The scale threshold scores that would indicate the need for investigation and subsequent support (whether education or remediation) will need confirming in other studies that include observation of consulting skills and other measures of performance.

It is possible that different patient experience questionnaires that address different attributes of consultations would produce different distributions of scores among GPs. Studies are required that compare the ability of different questionnaires to identify those doctors in need of investigation. Although the psychometric performance of questionnaires is important, in revalidation it is necessary to determine both the false-positive and the false-negative rates of identification of doctors in need of support or investigation. Only around one-sixth of patients added open comments to the questionnaire, and although these enabled some comparison between GPs with scores above and below the control limits, they did not provide specific guidance on what GPs should do differently to improve patient experience. Nevertheless, the study does provide evidence to inform a debate on how surveys of patient experience might be used in revalidation.

#### Comparison with existing literature

Similar studies have been performed for other patient questionnaires: the CARE instrument achieves reliability of 0.8 with 41 responders,<sup>17</sup> the GMC patient satisfaction questionnaire with 226 and more recently the same team has reported reliability of 0.7 with 35 responders.<sup>18</sup> Although evidence on the performance of questionnaires is therefore available, there is less evidence on how the findings should be used in revalidation. Pilot studies are underway, and although patient feedback was generally reported as valuable by participants in the pilots, its role was not considered in detail in the recent report.<sup>19</sup>

#### Implications for practice and research

This study demonstrates that patient surveys can be used to identify GPs who are outliers in terms of patient experience of their consultations. Using results from the CSQ, a small number of GPs were found

who received both low scores and a disproportionate amount of negative comments from patients. This helps justify the inclusion of patient satisfaction surveys in revalidation.<sup>20</sup> Some patients do express criticism of consultations with some GPs, and if patient experience is omitted from revalidation there is a risk that some doctors that patients regard as providing unsatisfactory consultations will go undetected and not receive help to improve. Also, evaluation of performance by patients draws attention to different aspects of performance than is elicited by evaluation of consultations by health professionals.<sup>21</sup> Other aspects of performance such as prescribing, referrals, or complaints may draw attention to a GP in need of investigation and subsequent support, but these performance measures may not always reflect consultation performance as perceived by patients. Unless evidence accumulates to show that GPs with poor consultation performance can be reliably identified by alternative performance measures, direct assessment of patient experience of consultations is required.

Before surveys of patient experience can be used in revalidation, however, a number of issues need to be resolved. First, most GPs will receive positive evaluations, and therefore have little to gain apart from affirmation from a system that is limited to identifying outliers. Consideration should be given to whether, and if so how, a process can be implemented to enable all GPs to reflect on patient experience and consider what they might do to improve it. Appraisal would be the natural setting, and this would help make revalidation a formative process for many GPs.

Secondly, it needs to be decided whether negative patient experience is sufficient on its own to trigger investigation, or whether investigation is triggered only when patient feedback is supplemented by other indications of poor performance. How much importance should be assigned by regulators to patient experience? This is likely to depend in part on the strength of dissatisfaction expressed by patients. For example, it may be argued that doctors who score below three or four standard deviations of the mean in all or most scales are in need of investigation, even if other measures of their performance are satisfactory. Doctors with slightly better patient feedback may be encouraged to attend consultation training courses as part of their continuing professional education activities, and to demonstrate improved performance in repeat surveys. The use of

---

#### Funding

No external funding was used in conducting this study.

#### Ethical approval

Ethics approval was received from the University of Leicester ethics committee (rb14-22fd).

#### Provenance

Freely submitted; externally peer reviewed.

#### Competing interests

Andrew Smith, as head of the primary care division of PatientDynamics, and Richard Baker as developer of CSQ, benefit financially from use of CSQ surveys.

#### Discuss this article

Contribute and read comments about this article on the Discussion Forum: <http://www.rcgp.org.uk/bjgp-discuss>

patient evaluations alone to trigger investigation is also likely to depend on the findings that accumulate during the early years of revalidation. If it becomes clear that some doctors who otherwise provide clinically appropriate and safe care have poor patient evaluations of their care, a question then follows as to whether this should be a matter for continuing education and appraisal processes rather than revalidation. On the other hand, if poor patient experience is found to be invariably associated with other indicators of poor performance, the pattern of poor performance across a range of indicators may become the usual trigger for investigation.

Thirdly, the threshold scores chosen to indicate the need, or possible need, for investigation and educational support have to be defined, and further research will be required to validate such threshold scores, as well as to evaluate the effectiveness of educational support and remediation. The control limits used in this study are somewhat arbitrary and do not constitute indications for investigation, education, or remediation. Finally, it may be necessary to consider standardising methods of administering patient surveys to minimise the theoretical risk of gaming, which could happen, for example, by inviting only those patients who are likely to be satisfied to take part in the survey.

## REFERENCES

1. General Medical Council. *Revalidation project initiation document*. London: General Medical Council, 2009. [http://www.gmc-uk.org/Project\\_Initiation\\_document.pdf\\_25397052.pdf](http://www.gmc-uk.org/Project_Initiation_document.pdf_25397052.pdf) [accessed 5 Jul 2011].
2. General Medical Council. *Revalidation — the way ahead. Consultation document*. London: General Medical Council, 2010.
3. General Medical Council. *Revalidation — the way ahead. Annex 3. GMC principles for criteria and key indicators for colleague and patient questionnaires in revalidation*. London: General Medical Council, 2010.
4. Royal College of General Practitioners. *RCGP guide to the revalidation of general practitioners, version 4*. London: Royal College of General Practitioners, 2010.
5. Royal College of General Practitioners. *Remediation for general practitioners. A RCGP policy document*. London: Royal College of General Practitioners, 2009. [http://www.rcgp.org.uk/pdf/Report\\_on\\_GP\\_Remediation.pdf](http://www.rcgp.org.uk/pdf/Report_on_GP_Remediation.pdf) [accessed 5 Jul 2011].
6. Campbell JL, Richard SH, Dickens A, *et al*. Assessing the professional performance of UK doctors: an evaluation of the utility of the General Medical Council patient and colleague questionnaires. *Qual Saf Health Care* 2008; **17(3)**: 187–193.
7. Lockyer J, Fidler H. *Comparison of patient satisfaction instruments designed for GPs in the UK. Report commissioned by the Royal College of General Practitioners*. London: Royal College of General Practitioners, 2009.
8. Baker R. Development of a questionnaire to assess patients' satisfaction with consultations in general practice. *Br J Gen Pract* 1990; **40(341)**: 487–490.
9. Baker R, Whitfield M. Measuring patient satisfaction: a test of construct validity. *Qual Health Care* 1992; **1(2)**: 104–109.
10. Kinnersley P, Stott N, Peters T, *et al*. A comparison of methods for measuring patient satisfaction with consultations in primary care. *Fam Pract* 1996; **13(1)**: 41–51.
11. McKinstry B, Colthart I, Walker J. Can doctors predict patients' satisfaction and enablement? A cross-sectional observational study. *Fam Pract* 2006; **23(2)**: 240–245.
12. PatientDynamics. *Consultation satisfaction questionnaire*. <http://www.csqsurvey.co.uk/welcome/index.html> [accessed 5 Jul 2011].
13. Streiner D, Norman G. *Health measurement scales: a practical guide to their development and use. 4th edn*. Oxford: Oxford Medical Publications, 2008.
14. G-String\_III. *A Windows wrapper for urGENOVA*. Hamilton: Program for Educational Research and Development, McMaster University, 2010. [http://fhspend.mcmaster.ca/g\\_string/index.html](http://fhspend.mcmaster.ca/g_string/index.html) [accessed 5 Jul 2011].
15. Feigenbaum AV. *Total quality control*. New York: McGraw Hill, 1991.
16. Healey JF. *Statistics. A tool for social research*. Belmont: Wadsworth Publishing, 1993.
17. Murphy DJ, Bruce DA, Mercer SW, Eva KW. The reliability of workplace-based assessment in postgraduate medical education and training: a national evaluation in general practice in the United Kingdom. *Adv Health Sci Educ Theory Pract* 2009; **14(2)**: 219–232.
18. Campbell JL, Wright C, Hill J, *et al*. Multi-source feedback for revalidation — report of large scale national pilot. Modelling patient and colleague feedback on doctor's professional performance. *SAPC ASM Proceedings* 2011; 79–80.
19. Department of Health/Revalidation Support Team. *Independent Evaluation of The Medical Revalidation Pathfinder Pilot. Final Report. Summary Report*. Durham University 2011. [http://www.revalidationsupport.nhs.uk/files/DE0270-00-DH-MedRev-Summary\\_Report\\_FINAL.pdf](http://www.revalidationsupport.nhs.uk/files/DE0270-00-DH-MedRev-Summary_Report_FINAL.pdf) [accessed 01 Sep 2011].
20. McKinley RK, Fraser RC, Baker R. Suggested model for directly assessing and enhancing clinical competence and performance in clinician revalidation. *BMJ* 2001; **322(7288)**: 712–715.
21. McKinley RK, Fraser RC, Baker RH, Riley RD. The relationship between measures of patient satisfaction and enablement and professional assessments of consultation competence. *Med Teach* 2004; **26(3)**: 223–228.