



PEARL

**Interpreting multisource feedback: online study of consensus and variation among GP appraisers**

Wright, Christine; Campbell, John; McGowan, Luke; Roberts, Martin J.; Jelley, Di; Chatterjee, Arunangsu

**Published in:**

British Journal of General Practice

**DOI:**

[10.3399/bjgp16X684373](https://doi.org/10.3399/bjgp16X684373)

**Publication date:**

2016

**Link:**

[Link to publication in PEARL](#)

**Citation for published version (APA):**

Wright, C., Campbell, J., McGowan, L., Roberts, M. J., Jelley, D., & Chatterjee, A. (2016). Interpreting multisource feedback: online study of consensus and variation among GP appraisers. *British Journal of General Practice*, 66(645), e277-e284. <https://doi.org/10.3399/bjgp16X684373>

All content in PEARL is protected by copyright law. Author manuscripts are made available in accordance with publisher policies. Wherever possible please cite the published version using the details provided on the item record or document. In the absence of an open licence (e.g. Creative Commons), permissions for further reuse of content should be sought from the publisher or author.

Wright et al  
Interpreting multisource feedback

This is a pre-proof version of an accepted article submitted to the British Journal of General Practice 2016 and published in vol. 66 iss. 645 pages e277 – e284.  
DOI: <http://dx.doi.org/10.3399/bjgp16X684373>

Please view the online version as edits and minor corrections may have been made.

1 **Interpreting multisource feedback: online study of consensus and variation among GP**  
2 **appraisers**

3 Christine Wright,<sup>1</sup> John Campbell,<sup>1</sup> Luke McGowan,<sup>2</sup> Martin J Roberts,<sup>3</sup> Di Jelley,<sup>4</sup> Arunangsu  
4 Chatterjee<sup>2,3</sup>

5 <sup>1</sup> Primary Care Research Group, University of Exeter Medical School, Exeter, Devon, UK.

6 <sup>2</sup> Technology Enhanced Learning for Medicine and Dentistry (TELMed) Team, Plymouth  
7 University Peninsula Schools of Medicine and Dentistry and Peninsula College of Medicine  
8 and Dentistry, Plymouth, Devon, UK.

9 <sup>3</sup> Collaboration for the Advancement of Medical Education Research and Assessment  
10 (CAMERA), Plymouth University Peninsula Schools of Medicine and Dentistry, Plymouth,  
11 Devon, UK.

12 <sup>4</sup> Health Education North East, Newcastle, UK.

13

14 **Correspondence should be addressed to:**

15 Professor John Campbell, Primary Care Research Group, University of Exeter Medical School,  
16 Smeall Building, St Luke's Campus, Magdalen Road, Exeter, Devon, United Kingdom, EX1

17 2LU; telephone: +44 (0)1392 722740; fax: +44 (0)1392 432223; e-mail:

18 [john.campbell@exeter.ac.uk](mailto:john.campbell@exeter.ac.uk)

19

20 **Word counts (for Original Articles):**

21 Title: 12 words

22 Abstract: 250 words

23 Manuscript (including suggested revisions): **2820 words**

24 Tables: 4                      Figures: 1                      Online files: 2

25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50

## Abstract

**Background:** GPs collect multisource feedback (MSF) about their professional practice and discuss it at appraisal. Appraisers use such information to identify concerns about a doctor's performance, and to guide the doctor's professional development plan (PDP).

**Aim:** To investigate whether GP appraisers detect variation in doctors' MSF results, and the degree of consensus in appraisers' interpretations of this information.

**Design and Setting:** Online study of GP appraisers in North-East England.

**Method:** GP appraisers were invited to review eight anonymised doctors' MSF reports, which represented different patterns of scores on the UK General Medical Council's Patient and Colleague Questionnaires. Participants provided a structured assessment of each doctor's report, and recommended actions for their PDP. Appraiser ratings of each report were summarised descriptively. An 'agreement score' was calculated for each appraiser to determine whether his/her assessments were more lenient than those of other participants.

**Results:** 101/146 appraisers (69%) assessed at least one report. The pattern of appraisers' ratings suggested they could detect variation in GPs' MSF results, and recommend reasonable actions for the doctors' PDP. Increasing appraiser age was associated with more favourable interpretations of MSF results.

**Conclusions:** Although preliminary, the finding of broad consensus amongst GP appraisers in their assessment of MSF reports should be reassuring for GPs, appraisers and employing organisations. However, if older appraisers are more lenient than younger appraisers in their interpretation of MSF and in the actions they suggest to their appraisees as a result, organisations need to consider what steps could be taken to address such differences.

**Key words:** General practitioners, appraiser, multisource feedback, revalidation, leniency

51 **How this fits in**

52

53

54 Doctors now collect and reflect on feedback from their patients and colleagues as part of the  
55 appraisal process. Little is known about how appraisers interpret multisource feedback (MSF)  
56 information.

57 This study explored GP appraisers' interpretations of a purposively-selected sample of MSF  
58 reports for eight doctors. The findings suggest appraisers can detect variation in GPs' MSF  
59 results and suggest appropriate action based on these, but appraisers may vary in the leniency  
60 / stringency of their interpretation of MSF information.

61

62

64 The UK General Medical Council (GMC) requires all practising doctors, over a five-year cycle,  
65 to collect supporting information to demonstrate adherence to the principles described in ‘*Good*  
66 *Medical Practice*’.<sup>1</sup> Doctors are expected to reflect on this information and discuss it as part of  
67 their appraisal process.<sup>2</sup> The supporting information includes multisource feedback (MSF) on  
68 the doctor’s practice obtained from colleagues and patients. MSF is viewed as a formative  
69 process, enabling individual doctors to identify where they may need to change their practice,  
70 and to plan their future professional development.<sup>2</sup>

71 A number of questionnaires are available to support the collection of MSF. The GMC has  
72 developed its own patient questionnaire (PQ) and colleague questionnaire (CQ), which assess  
73 various aspects of professional practice.<sup>3</sup> When feedback has been collated, each doctor is  
74 provided with a personalised report, summarising (for each core PQ/CQ item): the distribution  
75 of ratings of the doctor’s performance (5-point scales); a mean item percentage score;  
76 benchmark data derived from item percentage scores of other UK doctors; and the doctor’s  
77 self-assessment rating. Free text comments provided by the doctor, their patients and  
78 colleagues are also presented.

79 There is evidence that the GMC questionnaires are acceptable for use within appraisal to  
80 provide formative feedback on a doctor’s performance.<sup>3</sup> However, the resulting feedback can  
81 be complex and should be interpreted with caution.<sup>3;4</sup> Benchmark data are predominantly  
82 derived from volunteer doctor samples, and are markedly skewed towards positive views of  
83 performance. Thus an item score of 80-90% might still place a doctor in the lowest quartile  
84 when compared to their peers.<sup>3</sup> Furthermore, scores can be biased by factors associated with  
85 the individuals providing feedback or with the doctors themselves.<sup>3;4</sup>

86 Whilst the literature supports the use of MSF to improve practice,<sup>5-8</sup> a range of factors (relating  
87 to the individual doctor, their reaction to the feedback, and the availability of facilitation) may  
88 affect how a doctor uses the information to change their practice.<sup>6-8</sup>

89 GMC guidance recommends doctors discuss their MSF with an individual trained in providing  
90 feedback (such as their appraiser). Appraisers are expected to make “accurate and consistent  
91 judgements”<sup>9</sup> about supporting information to determine whether there are concerns about  
92 patient safety or the doctor’s conduct or performance. Resources have been developed to  
93 support appraisers in the wider process of revalidation<sup>10;11</sup> but these do not focus in detail on  
94 the interpretation of MSF. In one UK qualitative study,<sup>5</sup> appraisers of general practitioners  
95 (GPs) reported difficulty in interpreting benchmark information – that is, whether PQ/CQ item  
96 scores falling in the lowest quartile benchmark band are indicative of GP performance that  
97 should give cause for concern.

98 Little is known about the consistency of interpretation of MSF by GPs and their appraisers.  
99 However, research focusing on other ‘high-stakes’ performance-based assessments has  
100 observed examiner differences (‘hawk-dove effect’ or ‘stringency/leniency effect’)<sup>12-15</sup> that  
101 appear to be stable across time. In one UK study,<sup>14</sup> some examiners were observed to be  
102 more stringent (hawkish) in their assessment of candidates, and to require a higher level of  
103 performance for passing candidates than did other examiners. Whilst there was evidence that  
104 hawkishness correlated with examiner experience (number of candidates assessed) and ethnic  
105 origin, there was no evidence that it varied with examiner age or gender.<sup>14;16</sup> Other work in  
106 Canada suggests that individual examiners may be unaware of the extent of their  
107 stringency/hawkishness.<sup>17</sup>

## 108 **Study aims**

109 We piloted an online training resource to support the preparation of medical appraisers for their  
110 role in facilitating doctors’ reflection on MSF, within the context of UK appraisal and

111 revalidation. We aimed to: (1) assess appraisers' ability to detect variation in doctors' MSF  
112 scores; (2) explore the degree of consensus between appraisers with regard to their  
113 assessments of doctors' MSF results and actions they recommend; and (3) examine the  
114 variance between appraisers and identify potential demographic predictors of stringency or  
115 hawkishness in their interpretation of MSF.

## 116 **METHOD**

117 We designed and constructed an online training resource to provide GP appraisers with  
118 experience of interpreting MSF reports, and feedback on how their own interpretations  
119 compared to those of other appraisers. The design incorporated four clearly-labelled sections:  
120 background information about the project; instructions on using the resources; access to eight  
121 MSF reports (labelled 'A' to 'H'); and a feedback function allowing appraisers to compare their  
122 own assessment of each MSF report with assessments submitted by other appraisers.

123 Each MSF report summarised feedback for one GP in the format described above (see Online  
124 File A). Seven were real reports issued to UK GPs in earlier piloting of the GMC  
125 questionnaires.<sup>3</sup> At the end of that pilot work, standardised (Z) scores on the PQ and the CQ  
126 had been calculated for 402 doctors. A Z score below -1.96 was taken to indicate that the  
127 doctor's score fell in the lower tail of the distribution of doctor scores on the questionnaire (i.e.  
128 their score was statistically outlying). Based on the doctor's PQ and CQ Z scores, their report  
129 was categorised in one of four groups: (i) Neither PQ or CQ score statistically outlying; (ii) PQ  
130 score statistically outlying but CQ score not statistically outlying; (iii) CQ score statistically  
131 outlying but PQ score not statistically outlying; or (iv) PQ and CQ scores both statistically  
132 outlying.

133 Reports available on the online training resource were purposively selected to represent  
134 different patterns of PQ and CQ scores (Table 1), were anonymised and used with the doctors'  
135 explicit consent. Feedback indicative of poorer GP performance (group (iv) above) was rare in



136 the earlier pilot study;<sup>3</sup> therefore Report D was constructed to simulate such feedback.

137 Appraisers who assessed the reports were unaware of the doctors' actual Z scores.

138 Appraisers were asked to review each MSF report and complete a six-item online form. Three  
139 ordinal scale items evaluated the appraisers' interpretation of the doctor's MSF report: an  
140 overall assessment of the report (5-point scale: 'Excellent' to 'Unsatisfactory'); their level of  
141 concern about the GP's performance (4-point scale: 'Not at all concerned' to 'Extremely  
142 concerned'); and the acceptability of the GP's performance (4-point scale: 'Clearly acceptable'  
143 to 'Clearly unacceptable') based on the content of their MSF report. Three categorical items  
144 indicated the actions appraisers would discuss during the GP's appraisal: repeating the  
145 patient/colleague surveys; specific actions/training for the doctor's Personal Development Plan  
146 (PDP); and other possible actions. Respondents could also add free-text comments about the  
147 MSF report or their recommended actions.

148 The process was repeated for each MSF report in turn and appraisers could choose the order  
149 in which they assessed reports. Assessments could be completed over a number of sessions  
150 but could not be amended once submitted. After submitting an assessment, appraisers could  
151 access the feedback function to view a summary of other appraisers' assessments of the same  
152 report.

153 Preliminary user-testing of the online training resource was conducted (July-October 2012)  
154 with three GP appraisers to check the acceptability of the registration process, training  
155 exercise and supporting materials. Based on their feedback, changes were made to the  
156 training materials and web pages.

157 The revised training resource was made available to 235 GP appraisers from North-East  
158 England, across a series of waves (December 2012-November 2013). A panel of eight  
159 appraisers took part in the initial wave of recruitment and the panel's ratings of and comments  
160 about the constructed report (Report D) suggested this had face validity.

161 Appraisers were invited by the local appraisal lead to use the online resource as part of their  
162 continuing professional development. To register for an account, appraisers selected a user  
163 name/password and provided brief demographic information. Accounts were individually  
164 verified and activated by the researcher, after which appraisers could access the eight MSF  
165 reports. Up to two e-mail reminders were sent to non-responders.

166 Appraisers who assessed at least one MSF report were e-mailed a personalised record  
167 (December 2013) showing how their own assessments compared to those of other appraisers.  
168 The appraisal lead encouraged appraisers to reflect on the training exercise and their  
169 personalised record as part of their annual quality assurance review, and to discuss learning  
170 points in their local appraiser support group.

#### 171 **Statistical analysis**

172 We described the appraisers who used the online training resource in terms of their gender,  
173 age, ethnic origin, region of primary medical qualification (PMQ), and appraisal experience.

174 The characteristics of appraisers who assessed at least one MSF report ('participants') were  
175 compared to those who registered but did not assess any reports ('non-participants') using Chi-  
176 Square ( $\chi^2$ ) tests for categorical variables and Mann-Whitney (U) tests for continuous  
177 variables.

178 For each MSF report, we described the frequency distribution of responses on the six  
179 assessment items and, for the three ordinal scale items (overall assessment, concerns, and  
180 acceptability), we calculated the mode, mean and standard deviation (SD) of the ratings.

181 For appraisers who assessed all eight MSF reports, an 'agreement score' was calculated by  
182 summing the differences between their overall assessment rating and the modal rating of all  
183 appraisers on each of the reports. Negative agreement scores were indicative of hawk-like

184 tendencies (on average rating reports less favourably than peers), while positive scores were  
185 indicative of dove-like tendencies (on average rating reports more favourably than peers).

186 We described the distribution of these agreement scores and conducted an analysis of  
187 variance (ANOVA) to explore the effects of gender, age (4 categories), ethnicity (2 categories:  
188 White, Other), PMQ (2 categories: UK, Other) and years of appraiser experience as predictors  
189 of hawk-like/dove-like tendencies. *P* values of less than 0.05 were regarded as statistically  
190 significant.

## 191 **RESULTS**

### 192 **Participants**

193 In total, 146/235 (62%) appraisers registered to use the online training resource, of whom  
194 101/146 (69%) assessed at least one MSF report and 86/146 (59%) assessed all eight reports.

195 Table 2 describes the characteristics of participating and non-participating appraisers. Non-  
196 participants were more likely than participants to be male ( $P=0.03$ ) and from non-White ethnic  
197 groups ( $P=0.02$ ). However, the two groups were similar in terms of age, region of PMQ, and  
198 experience as an appraiser (all  $P > 0.05$ ).

### 199 **MSF report assessments**

200 Participants' overall assessments of Reports A to H are summarised in Figure 1. Detailed data  
201 on the distribution of their responses on all three evaluative scales appear in Online File B. The  
202 pattern of modal ratings (Table 3) suggests appraisers were broadly able to detect variation in  
203 GPs' MSF results.

204 Most reports (7/8) received a modal overall assessment of 'Satisfactory' or higher; only Report  
205 D had a mode assessment of 'Borderline'. Mean concern ratings about GP performance were  
206 highest for Reports D, C, G and B. More than half of appraisers reported 'significant concerns'  
207 about the GP performance reflected in Report D (statistically outlying on PQ and CQ) and in

208 Report C (outlying on CQ only). A similar proportion reported 'minor concerns' about the  
209 doctors' performance reflected in Reports G (outlying on CQ) and B (outlying on PQ). Doctor  
210 performance was rated by the majority of appraisers as being 'clearly acceptable' or 'probably  
211 acceptable' for all reports. However, one quarter to one third of appraisers rated the  
212 performance of the doctors assessed in Reports C and D as 'probably unacceptable' or 'clearly  
213 unacceptable'.

214 Given the formative purpose of MSF, appraisers appeared to recommend reasonable actions  
215 (Table 4) in the form of repeating one or both surveys, and the inclusion of training in the  
216 doctor's PDP. Additional actions were suggested by one-quarter of appraisers in response to  
217 Report D, most commonly recommending the appraiser sought advice from a GP tutor or  
218 appraisal supervisor, discussed mental wellbeing/stress management with the doctor, and  
219 explored the doctor's insight into their communication skills. Only a minority of appraisers  
220 recommended referring GPs to their Responsible Officer for further review (1%, 5% and 7% for  
221 Reports B, C and D respectively).

## 222 **Hawk-dove effects**

223 Agreement scores, reflecting the difference between individual appraisers' assessment ratings  
224 and the modal rating for the eight MSF reports, ranged from -7 to +7 (mean agreement score  
225 0.49; SD=3.01). An agreement score of -7 would indicate a hawk-like appraiser who might, for  
226 example, have rated seven of the eight reports at one point below the modal rating (e.g.  
227 'Borderline' rather than 'Satisfactory') and agreed with the modal score on just one report.  
228 Conversely, an agreement score of +7 would indicate a dove-like appraiser who might also  
229 have given the modal rating on one report but rated the other seven reports at one point higher  
230 than the mode. Hawk-like tendencies were more common (44/86, 51% appraisers with a  
231 negative agreement score) than dove-like tendencies (29/86, 34% appraisers with a positive

232 agreement score). Despite this, the mean agreement score was positive, indicating that the  
233 dove-like raters tended to deviate more from the modal rating than did the hawk-like raters.  
234 Age was a significant predictor of hawk/dovelike tendencies, with older appraisers rating the  
235 MSF reports more favourably than younger appraisers ( $B=0.129$ ,  $P=0.01$ ). Gender, ethnic  
236 origin, PMQ and years as an appraiser were not, however, significant predictors of  
237 hawk/dovelike tendencies.

## 238 **DISCUSSION**

239 Despite the complexity of information in the featured MSF reports, appraisers' assessments  
240 suggested they could detect variations in MSF score patterns. For each report, there was  
241 broad consensus about the level of concern and acceptability of the GP's performance (based  
242 on the information in their MSF report) and about actions that might be discussed in the  
243 appraisal meeting. However, appraisers varied in their tendency to be more stringent or lenient  
244 in their assessment of MSF reports relative to their peers. In particular, there was some  
245 evidence that older appraisers may be more lenient than younger appraisers in this regard.

### 246 **Comparison with existing literature**

247 Our observation that individual appraisers may vary in the leniency of their assessments of  
248 MSF reports is in line with hawk-dove effects observed in relation to other practice-based  
249 assessments.<sup>12-17</sup> Previous research has identified demographic characteristics of assessors  
250 that may be associated with variations in leniency (such as ethnic origin and experience).<sup>14;16</sup>  
251 Our study has identified appraiser age, but not length of experience as an appraiser, as a  
252 potential predictor of greater leniency in interpreting MSF reports. Appraisers from non-White  
253 ethnic backgrounds were under-represented in our sample and this may account for the  
254 absence of an observed effect of ethnic origin on leniency in our study.

255

## 256 **Strengths and limitations**

257 Seven of the eight MSF reports had been issued to practising GPs,<sup>3</sup> and therefore appraisers  
258 assessed realistic MSF information. The design of the online resource meant that appraisers  
259 could review reports over several sessions to fit around their work schedule. However,  
260 appraisers' judgements about a doctor's performance were made solely on the basis of an  
261 MSF report, without access to the doctor's other supporting information, or knowledge of the  
262 doctor's reaction to their feedback, which would occur in a real appraisal context.

263 A number of MSF tools that include different items, scales and reporting formats are available  
264 to doctors. Our study focused only on the interpretation of MSF reports derived from the GMC  
265 patient and colleague questionnaires.

266 GP appraisers were drawn from one region of the UK, which limits generalisability to other  
267 regions and contexts in which MSF is used. Relatively small numbers of appraisers  
268 participated (N=101), and participants were more likely to be female and from white ethnic  
269 backgrounds. Our findings should therefore be regarded as preliminary and interpreted  
270 cautiously until replicated with other appraiser samples.

271 With regard to hawk-dove effects in interpreting MSF, we collected limited demographic  
272 information about participating appraisers and other factors not addressed in this study may be  
273 associated with the observed variation in stringency/leniency.

## 274 **Implications for research and/or practice**

275 Our study suggests appraisers can detect variation in the pattern of GPs' MSF scores and  
276 recommend appropriate actions based on a review of complex MSF information. Furthermore,  
277 as a group, our appraisers were reasonably consistent in their interpretations of each doctor's  
278 MSF results. This observation should be reassuring for GPs and appraisers, as well as for

279 appraisal leads, Responsible Officers and Designated Bodies<sup>11</sup> who have responsibility for  
280 quality assurance of appraisal processes.

281 Individual differences in leniency were observed in appraisers' interpretations of MSF, which  
282 may be linked to the appraiser's age. GPs' experiences of reflecting and acting on MSF within  
283 their appraisal may therefore vary according to the age of their appraiser. The extent to which  
284 this proves problematic in real-life practice is yet to be established. Similarly, the need for  
285 organisations to take steps to attenuate appraiser differences in leniency around MSF requires  
286 further consideration. This might include the use of training packages utilising standardised  
287 reports such as described in this study. Future development work could evaluate appraisers'  
288 views of our online training resource and determine how it might be improved by seeking  
289 feedback from appraisers who assess all eight MSF reports as well as those who assess fewer  
290 reports.

291 Research employing qualitative or cognitive interviewing methods might explore how  
292 appraisers arrive at judgements about a doctor's performance based on MSF reports, and  
293 which aspects of the available MSF information influence their interpretations. Further study of  
294 hawk-dove effects in this context could identify why such differences exist, how appraisers  
295 view their own level of stringency, and whether these effects change after using the training  
296 resource or change with increasing experience of interpreting MSF in the context of 'real-world'  
297 appraisal.

298

299 **References**

- 300 1. General Medical Council. Good Medical Practice. Manchester: UK General Medical  
301 Council, 2013.
- 302 2. General Medical Council. Supporting information for appraisal and revalidation.  
303 Manchester: General Medical Council, 2012.
- 304 3. Wright C, Richards SH, Hill JJ, Roberts MJ, Norman GR, Greco M *et al*. Multisource  
305 feedback in evaluating the performance of doctors: The example of the UK General  
306 Medical Council Patient and Colleague Questionnaires. *Acad Med* 2012;**87**:1668-78.
- 307 4. Campbell JL, Roberts M, Wright C, Hill J, Greco M, Taylor M *et al*. Factors associated  
308 with variability in the assessment of UK doctors' professionalism: analysis of survey  
309 results. *BMJ* 2011;**343**:d6212.
- 310 5. Hill JJ, Asprey A, Richards SH, Campbell JL. Multisource feedback questionnaires in  
311 appraisal and for revalidation: a qualitative study in UK general practice. *Br J Gen Pract*  
312 2012;**62**:e314-e321.
- 313 6. Miller A, Archer J. Impact of workplace based assessment on doctors' education and  
314 performance: a systematic review. *BMJ* 2010;**341**: c5064.
- 315 7. Sargeant J, Mann K, Ferrier S. Exploring family physicians' reactions to multisource  
316 feedback: perceptions of credibility and usefulness. *Medical Education* 2005;**39**:497-504.
- 317 8. Sargeant J, Mann K, Sinclair D, van der Vleuten C, Metsemakers J. Understanding the  
318 influence of emotions and reflection upon multi-source feedback acceptance and use.  
319 *Advances in Health Sciences Education* 2008;**13**:275-88.
- 320 9. NHS Revalidation Support Team. Quality Assurance of Medical Appraisers: Recruitment,  
321 training, support and review of medical appraisers in England. London: NHS Revalidation  
322 Support Team, 2013.
- 323 10. NHS Revalidation Support Team. Medical Appraisal Guide: A guide to medical appraisal  
324 for revalidation in England. London: NHS Revalidation Support Team, 2013.
- 325 11. Royal College of General Practitioners. The Principles of GP Appraisal for Revalidation.  
326 London: Royal College of General Practitioners, 2014.
- 327 12. Bartman I, Roy M, and Smee S. Catching the Hawks and Doves: A Method for Identifying  
328 Extreme Examiners on Objective Structured Clinical Examinations (Technical Report).  
329 Ottawa: Medical Council of Canada, 2011.
- 330 13. Harasym PH, Woloschuck W, Cuning L. Undesired variance due to examiner  
331 stringency/leniency effect in communication skill scores in OSCEs. *Advances in Health*  
332 *Sciences Education* 2008;**13**:617-32.
- 333 14. McManus IC, Thompson M, Mollon J. Assessment of examiner leniency and stringency  
334 ('hawk-dove effect') in the MRCP (UK) clinical examination (PACES) using multi-facet  
335 Rasch modelling. *BMC Medical Education* 2006;**6**:42.
- 336 15. Roberts C, Rothnie I, Zoanetti N. Should candidate scores be adjusted for interviewer  
337 stringency or leniency in multiple mini-interviews? *Medical Education* 2010;**44**:690-8.



- 338 16. McManus IC, Elder AT, Dacre J. Investigating possible ethnicity and sex bias in clinical  
339 examiners: an analysis of data from the MRCP(UK) PACES and nPACES examinations.  
340 *BMC Medical Education* 2013;**13**:103.
- 341 17. Bartman I, Smee S, Roy M. A method for identifying extreme OSCE examiners. *Clinical*  
342 *Teacher* 2013;**10**:27-31.  
343  
344
- 345
- 346

347  
348 **Acknowledgements:** The authors thank the appraisers who piloted and helped to refine the  
349 online training resource materials in the various development phases. They also thank Mr Zac  
350 Gribble and Ms Sally Holden (formerly of the E-Learning Support Team, Peninsula College of  
351 Medicine and Dentistry), who provided technical expertise during the development of the initial  
352 prototype of the online training resource.

353 **Funding/support:** The work was supported by a research grant from the General Medical  
354 Council (GMC) and funds allocated by Health Education North East for appraiser training.

355 **Other disclosures:** Professor Campbell was an advisor to the GMC during the development  
356 of the GMC patient and colleague questionnaires (2005-2011) and received only direct costs  
357 associated with presentation of that work. All other authors have no conflicts of interest to  
358 declare.

359 **Ethical approval:** The project represented the development and piloting of a local training  
360 resource for NHS appraisers and therefore did not require NHS ethics approval.

361 **Previous presentations:** A summary of this work was presented as an 'elevator pitch' at the  
362 Society for Academic Primary Care (SAPC) National Meeting in Edinburgh, Scotland (11 July  
363 2014).

364

365

366

367

368 **Author contributions:**

369 Dr Wright developed materials for the online training resource, managed the recruitment and  
370 registration of appraisers, oversaw the online data collection, analysed the resulting data,  
371 drafted the manuscript, and edited later versions.

372 Professor Campbell jointly conceived the idea for the project, contributed to the development of  
373 the online training materials, supervised the project, reviewed the data analysis, and  
374 commented on all versions of the manuscript.

375 Mr McGowan provided technical advice and support in relation to the development and  
376 revision of the online training resource, managed specific technical problems that arose, and  
377 commented on all versions of the manuscript.

378 Mr Roberts provided general statistical advice and conducted the analysis to explore predictors  
379 of hawk-dove like tendencies among appraisers, and commented on all versions of the  
380 manuscript.

381 Dr Jelley jointly conceived the idea for the project, contributed to the development of the online  
382 training materials, helped to identify and approach appraisers in the North East of England,  
383 and commented on all versions of the manuscript.

384 Dr Chatterjee supervised the technical aspects of the online training resource development  
385 and commented on all versions of the manuscript.

386

387

388 **List of Tables**

- 389 Table 1: Overview of multisource feedback reports available for review by appraisers  
390  
391 Table 2: Characteristics of participating and non-participating appraisers  
  
392 Table 3: Appraisers' modal evaluations (N, % participants) of feedback reports A to H with  
393 reports grouped by the pattern of doctor's feedback scores  
394  
395 Table 4: Appraisers' suggested actions for feedback reports A to H – distribution of  
396 responses

397

398 **List of Figures**

- 399 Figure 1: Distribution of appraisers' evaluations of feedback reports A to H (Overall  
400 assessment ratings)

401

402 **Potential online files**

403

- 404 Online File A: Example multisource feedback report (Report D) featured in online training  
405 resource.

- 406 Online File B: Appraisers' evaluations of feedback reports A to H – distribution and mean  
407 ratings (for overall assessment of report; concerns about doctor's  
408 performance; and acceptability of doctor's performance)  
409

410

**Table 1: Overview of multisource feedback reports available for review by appraisers**

Report	Outlying PQ score? <sup>†</sup>	Outlying CQ score? <sup>†</sup>	N (%) PQ item scores in lower quartile band <sup>‡</sup>	N (%) PQ item scores in upper quartile band <sup>¶</sup>	N (%) CQ item scores in lower quartile band <sup>‡</sup>	N (%) CQ item scores in upper quartile band <sup>¶</sup>	Areas of concern highlighted in patient (P) or colleague (C) free text comments
A	No (Z = 0.29)	No (Z = 0.10)	1/9 (11%)	3/9 (33%)	3/18 (17%)	3/18 (17%)	None – all comments positive
B	<b>Yes</b> <b>(Z = -2.02)</b>	No (Z = -0.72)	8/9 (89%)	0/9 (0%)	4/18 (22%)	3/18 (17%)	None – all comments positive
C	No (Z = -1.41)	<b>Yes</b> <b>(Z = -2.45)</b>	7/9 (78%)	0/9 (0%)	18/18 (100%)	0/18 (0%)	Record keeping; prescribing (C)
D	<b>Yes</b> <b>(Z = -4.46)</b>	<b>Yes</b> <b>(Z = -2.19)</b>	5/9 (55%)	0/9 (0%)	11/18 (61%)	0/18 (0%)	Interpersonal skills (P,C)
E	No (Z = 0.12)	No (Z = 0.37)	0/9 (0%)	0/9 (0%)	1/18 (6%)	6/18 (33%)	None – all comments positive
F	No (Z = -1.63)	No (Z = -0.83)	9/9 (100%)	0/9 (0%)	9/18 (50%)	3/18 (17%)	Record keeping; aloofness (C)
G	No (Z = -1.08)	<b>Yes</b> <b>(Z = -1.93)</b>	6/9 (67%)	0/9 (0%)	13/18 (72%)	0/18 (0%)	Managing time/workload (C)
H	No (Z = 0.69)	No (Z = 0.33)	0/9 (0%)	7/9 (78%)	1/18 (6%)	3/18 (17%)	None – all comments positive

<sup>†</sup> Outlying Patient Questionnaire (PQ) or Colleague Questionnaire (CQ) overall scores were defined as those lying more than 1.96 standard deviations below the mean PQ or CQ overall score (standardised Z score < -1.96) calculated for all doctors who participated in GMC questionnaire pilot work.<sup>3</sup>

<sup>‡</sup> Number of PQ or CQ core items where the doctor's score fell in the lowest 25% of item scores achieved by doctors who participated in GMC questionnaire pilot work.<sup>3</sup>

<sup>¶</sup> Number of PQ or CQ core items where the doctor's score fell in the highest 25% of item scores achieved by doctors who participated in previous pilot work.<sup>3</sup>

**Table 2: Characteristics of participating and non-participating appraisers**

		Participating appraisers <sup>†</sup> (N = 101)	Non-participating appraisers <sup>‡</sup> (N = 45)	Statistical tests
Gender	Male	52 (51%)	32 (71%)	$\chi^2$ (df1) = 4.908; <b>P = 0.03</b>
	Female	49 (48%)	13 (29%)	
Age group	30 to 39 years	10 (10%)	3 (7%)	$\chi^2$ (df3) = 3.931; P = 0.27
	40 to 49 years	35 (35%)	10 (22%)	
	50 to 59 years	43 (43%)	26 (58%)	
	60 years or over	9 (9%)	6 (13%)	
	<i>Missing</i>	4 (4%)	0 (0%)	
Ethnic group	White	86 (85%)	33 (73%)	$\chi^2$ (df1*) = 6.144; <b>P = 0.02</b>
	Mixed	0 (0%)	1 (2%)	
	Asian or Asian British	9 (9%)	10 (22%)	
	Chinese or Other Group	1 (1%)	1 (2%)	
	<i>Missing</i>	5 (5%)	0 (0%)	
Region of primary medical qualification (PMQ)	United Kingdom	86 (85%)	39 (87%)	$\chi^2$ (df1*) = 0.116; P = 0.78
	European Economic Area	2 (2%)	2 (4%)	
	South Asia	7 (7%)	4 (9%)	
	Other	2 (2%)	0 (0%)	
	<i>Missing</i>	4 (4%)	0 (0%)	
Length of experience as medical appraiser (in years)		Median = 7.0 LQ = 3; UQ = 10 <sup>‡</sup> (Range 0-38)	Median = 8.0 LQ = 3; UQ = 10 <sup>‡</sup> (Range 0-15)	(Mann-Whitney U = 2039.50, P = 0.47).

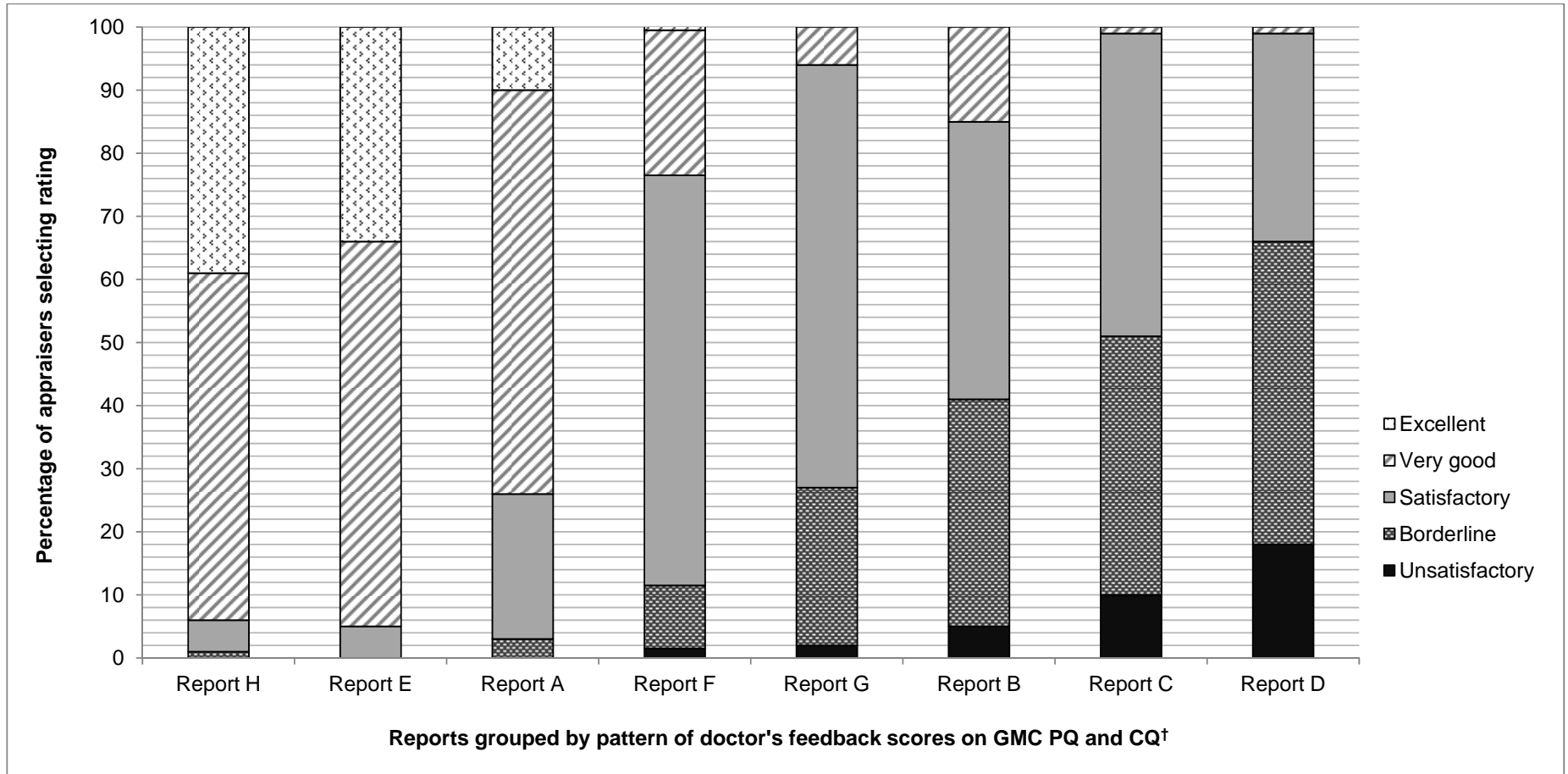
<sup>†</sup> All participating appraisers who assessed at least one MSF report (excluding 3 appraisers who participated in preliminary user-testing in July-October 2012).

<sup>‡</sup> LQ = lower quartile; UQ = upper quartile.

<sup>¶</sup> Non-participating appraisers registered to use the online training resource but submitted no MSF report assessments.

\* Chi square test for ethnic group compared White vs. Other ethnic groups; Chi square test for region of PMQ compared UK vs. Other PMQ regions.

**Figure 1: Distribution of appraisers' evaluations of feedback reports A to H (Overall assessment ratings)**



† Notes on grouping of reports: (i) Reports A, E, H and F: Doctors' scores on the PQ and the CQ were not statistically outlying in previous piloting of the GMC questionnaires (Wright *et al*, 2012). (ii) Reports G, B, and C: Doctors' scores on one questionnaire (either the PQ or the CQ) were statistically outlying in previous piloting of the GMC questionnaires. (iii) Report D: Doctor's scores on both questionnaires (the PQ and the CQ) would have been statistically outlying in previous piloting of the GMC questionnaires.

**Table 3: Appraisers' modal evaluations (N, % participants) of feedback reports A to H with reports grouped by the pattern of doctor's feedback scores**

Report	Pattern of doctor's feedback scores	Evaluative item		
		Overall assessment of report	Level of concern about doctor's performance	Acceptability of doctor's performance
<b>A</b>	Fell within 'normal distribution' of scores on PQ <u>and</u> on CQ	Very good (64/100; 64%)	Not at all (58/100; 58%)	Clearly acceptable (73/100; 73%)
<b>E</b>		Very good (54/88; 61%)	Not at all (76/88; 86%)	Clearly acceptable (86/88; 98%)
<b>H</b>		Very good (48/87; 55%)	Not at all (64/87; 74%)	Clearly acceptable (80/87; 92%)
<b>F</b>		Satisfactory (58/89; 65%)	Minor only (62/89; 70%)	Probably acceptable (47/89; 53%)
<b>B</b>	Outlier <sup>†</sup> on PQ only	Satisfactory (42/95; 44%)	Minor only (44/95; 46%)	Probably acceptable (58/95; 61%)
<b>C</b>	Outlier <sup>†</sup> on CQ only	Satisfactory (45/93; 48%)	Significant (48/93; 52%)	Probably acceptable (63/93; 68%)
<b>G</b>	Outlier <sup>†</sup> on CQ only	Satisfactory (58/87; 67%)	Minor only (55/87; 63%)	Probably acceptable (63/87; 72%)
<b>D</b>	Outlier <sup>†</sup> on PQ <u>and</u> on CQ	Borderline (44/91; 48%)	Significant (62/91; 68%)	Probably acceptable (57/91; 63%)

<sup>†</sup> Outlying Patient Questionnaire (PQ) or Colleague Questionnaire (CQ) overall scores were more than 1.96 standard deviations below the mean PQ or CQ overall score (standardised Z score  $\leq -1.96$ ) calculated for all doctors who participated in GMC questionnaire pilot work.<sup>3</sup>



**Table 4: Appraisers' suggested actions for feedback reports A to H – distribution of responses**

	Report A N=100	Report B N=95	Report C N=93	Report D N=91	Report E N=88	Report F N=89	Report G N=87	Report H N=87
<b>Repeating patient and colleague surveys</b>								
No need to repeat either survey	<b>90 (90%)</b>	<b>45 (47%)</b>	<b>43 (46%)</b>	17 (19%)	<b>87 (99%)</b>	<b>57 (64%)</b>	<b>46 (53%)</b>	<b>85 (98%)</b>
Repeat the patient survey only	0 (0%)	<b>45 (47%)</b>	3 (3%)	33 (36%)	1 (1%)	12 (13%)	6 (7%)	2 (2%)
Repeat the colleague survey only	7 (7%)	0 (0%)	17 (18%)	2 (2%)	0 (0%)	2 (2%)	20 (23%)	0 (0%)
Repeat both surveys	3 (3%)	5 (5%)	30 (32%)	<b>39 (43%)</b>	0 (0%)	18 (20%)	15 (17%)	0 (0%)
<b>Personal Development Plan (PDP) actions</b>								
No specific PDP action(s) or training	43 (43%)	12 (13%)	1 (1%)	1 (1%)	<b>65 (74%)</b>	18 (20%)	5 (6%)	<b>63 (72%)</b>
Encourage to include training in PDP	<b>55 (55%)</b>	<b>57 (60%)</b>	<b>49 (53%)</b>	40 (44%)	23 (26%)	<b>61 (69%)</b>	<b>62 (71%)</b>	24 (28%)
Mandate to include training in PDP	2 (2%)	26 (27%)	43 (46%)	<b>50 (55%)</b>	0 (0%)	10 (11%)	20 (23%)	0 (0%)
<b>Other recommended action(s) †</b>								
No other action needed	<b>70 (70%)</b>	32 (34%)	7 (8%)	2 (2%)	<b>77 (88%)</b>	<b>42 (47%)</b>	24 (28%)	<b>63 (72%)</b>
Review PDP actions at next appraisal	30 (30%)	<b>60 (63%)</b>	<b>80 (86%)</b>	<b>73 (80%)</b>	8 (9%)	<b>42 (47%)</b>	<b>60 (69%)</b>	17 (20%)
Refer to the Responsible Officer	0 (0%)	1 (1%)	4 (4%)	6 (7%)	0 (0%)	0 (0%)	1 (1%)	0 (0%)
Other action recommended	0 (0%)	8 (8%)	13 (14%)	25 (27%)	3 (3%)	9 (10%)	10 (11%)	7 (8%)

Note: Emboldened figures represent modal response(s). † Percentages may add up to more than 100% because appraisers could select more than one action.

## **Doctor's Revalidation Online Support Project**

### **Peninsula multisource feedback benchmarking and standard setting application**

# **Report D**

**Please review the following feedback report and complete the assessment form for 'Report D' which you will find on the project website.**

**On the assessment form, you will be asked to:**

- Provide an overall grading of this doctor's report;
- Rate the level of concern you have about this doctor's performance;
- Rate the extent to which you believe this doctor's performance is acceptable;
- Indicate what action you would recommend, based on the feedback provided in this doctor's report.
- Provide any additional comments about this doctor's feedback, and/or the reasons for your recommended action.

## Report Contents

### Introduction

#### Patient Feedback

Patient demographics (table 1.1)	P1
Evaluation question ratings and scores	
Distribution and frequency of ratings (table 1.2)	P2
Mean percentage scores and benchmarks (table 1.3)	P3
Patient comments	P4

#### Colleague Feedback

Colleague demographics (table 2.1)	C1
Evaluation question ratings and scores	
Distribution and frequency of ratings (table 2.2)	C2
Mean percentage scores and benchmarks (table 2.3)	C3
Colleague comments	C4

#### Self Assessment

Comparison of self assessed scores with patient and colleague scores (table 3.1)	S1
Personal comment	S2

#### Setting-specific benchmarks

Background information	B1
Patient setting-specific benchmarks (table 4.1)	B2
Colleague setting-specific benchmarks (table 4.2)	B3

## Supporting documents

Please see 'Background Information' document on the appraiser training website for:

Details of score calculation

Explanation of quartiles

Sample patient and colleague questionnaires

# Patient Feedback

## Your patient demographics

Your patient feedback is based on responses from 42 patients with the following characteristics:

Table 1.1: Gender

Female	23	55%
Male	19	45%
Not reported	0	0%

### Age

15-20	3	7%
21-40	10	24%
41-60	12	28%
Over 60	15	36%
Not reported	2	5%

Number and percentage of responses by question (percentage of responses may not add up to 100% due to rounding). 'Not reported' includes respondents who (a) did not answer this question; or (b) did answer but formed part of a category in which there were less than 3 respondents (and thereby are not displayed in order to protect personal anonymity).

## Evaluation question ratings and scores

Table 1.2: Distribution and frequency of ratings (Q4 - Q8)

	Poor	Less than satisfactory	Satisfactory	Good	Very good	Does not apply	Blank (missing)	Spoilt
Q4a Being polite	2	2	20	10	6	0	1	1
Q4b Making you feel at ease	1	2	24	10	4	0	1	0
Q4c Listening to you	1	2	20	14	5	0	0	0
Q4d Assessing your medical condition	0	0	1	8	33	0	0	0
Q4e Explaining your condition and treatment	0	0	2	10	24	6	0	0
Q4f Involving you in decisions	0	1	8	11	15	6	1	0
Q4g Providing or arranging treatment for you	0	0	0	9	26	7	0	0

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree	Does not apply	Blank (missing)	Spoilt
Q5a Confidentiality of information	0	0	0	14	26	0	2	0
Q5b Doctor is honest and trustworthy	0	0	0	12	28	0	2	0




	Yes	No	Spoilt	Blank (missing)
Q6 I am confident about this doctor's ability to provide care	38	0	0	4
Q7 I would be completely happy to see this doctor again	30	6	0	6

Blank, 'does not apply' and spoilt responses are not included in the score calculation as presented in table 1.3  
 - insufficient responses to display data

## Evaluation question ratings and scores

Table 1.3: Mean percentage scores and benchmarks (Q4,Q5)

	Your mean score (%)	Benchmark data (%) *				
		Min	Lower quartile	Median	Upper quartile	Max
Q4a Being polite	60	70	96	98	99	100
Q4b Making you feel at ease	59	69	94	97	98	100
Q4c Listening to you	62	61	94	97	98	100
Q4d Assessing your medical condition	94	68	93	96	98	100
Q4e Explaining your condition and treatment	90	65	93	95	98	100
Q4f Involving you in decisions	79	67	92	95	97	100
Q4g Providing or arranging treatment for you	94	68	93	96	98	100
Q5a Confidentiality of information	91	59	90	93	95	100
Q5b Doctor is honest and trustworthy	93	65	91	94	96	100

	Your mean score for this question falls in the highest 25% of all means
	Your mean score for this question falls in the middle 50% of all means
	Your mean score for this question falls in the lowest 25% of all means

- insufficient number of responses to generate score

See score explanation for percentage score calculation and quartile information

### \* Important notes about the benchmark data

The above benchmarks are based on all data that is available for the doctors who participated in our pilot work up to 2 July 2009. To calculate the benchmarks for each questionnaire item, we have included data only for doctors who had at least six valid responses returned for that particular item. We will update these benchmarks at regular intervals as the number of doctors undertaking these surveys increases.

Please bear in mind the following points when interpreting the above table:

**Purpose:** These benchmarks are provided to give you a sense of how you are performing in relation to other doctors who have completed the GMC surveys. They are not intended to imply any 'minimum standard' that doctors are expected to achieve for the purposes of revalidation.

**Sample size:** Of the doctors who have agreed to take part in the pilot work so far, 935 have completed a patient survey and 942 have completed a colleague survey. The benchmark data contained in this report is derived only from the doctors who returned sufficient numbers of patient or colleague responses. To date, 560 doctors have returned sufficient patient feedback and 620 doctors have returned sufficient colleague feedback.

**Voluntary participation:** It is not mandatory for any doctor to undertake these surveys. Therefore the benchmarks are based on a volunteer sample of doctors. As such, the benchmarks may be higher than might be expected if all doctors had contributed data.

**Range of practice:** The doctors who have contributed to the benchmark data work or provide care in a variety of settings and specialties. The above benchmarks relate to the whole sample of doctors who have contributed data, irrespective of their setting or specialty.

## Patient comments

From the free text component of the questionnaire

All comments have been included in their entirety but details which could identify a specific practitioner, practice or patient have been removed to ensure anonymity.

Please add any other comments you want to make about this doctor

I have every confidence in this doctor's ability/knowledge – but I think his 'bedside manner' could be improved.

I have known this doctor for a long time and I have always found him to be honest and helpful. At first I was rather nervous of him, but I have relaxed over time.

I have always been treated well.

Thorough and quick, but perhaps too busy to engage fully with patient.

Not as good a communicator as some of the other doctors at this Practice. The doctor could try to smile more because he doesn't make people feel at ease.

The doctor is a little dismissive when you mention your worries or concerns, but doesn't seem to understand how these might make you feel.

No complaints - good standard of care.

I am sure this doctor is very knowledgeable and experienced, but I personally would have got more from the consultation if he had a warmer approach.

The doctor needs to manage his time better. He was running very late but made no apology for this. Sometimes his manner can appear abrupt.

I have been with this doctor for a number of years and I am quite happy with the treatment I am given. I've got no complaints. Everything is satisfactory.

Got directly to the point, very pragmatic.

The doctor was professional and efficient. However wasn't given the chance to ask questions or raise my concerns. That would be my only issue for him to improve on.

I was satisfied with the care and information given by this doctor.

# Colleague Feedback

## Your colleague demographics

Your colleague feedback is based on responses from 18 colleagues with the following characteristics:

Table 2.1: Gender

Female	10	56%
Male	8	44%

### Age

30 - 49	9	50%
50 or over	9	50%

### Professional role

Doctor	9	50%
Other healthcare professional	7	39%
Not reported	2	11%

Number and percentage of responses by question (percentage of responses may not add up to 100% due to rounding). 'Not reported' includes respondents who (a) did not answer this question; or (b) did answer but formed part of a category in which there were less than 3 respondents (and thereby are not displayed in order to protect personal anonymity).



## Evaluation question ratings and scores

Table 2.2: Distribution and frequency of ratings (Q1 - Q19)

	Poor	Less than satisfactory	Satisfactory	Good	Very good	Don't know	Blank (missing)	Spoilt
Q1 Clinical knowledge	0	0	0	5	12	1	0	0
Q2 Diagnosis	0	0	0	5	12	1	0	0
Q3 Clinical decision making	0	0	1	4	10	3	0	0
Q4 Treatment including practical procedures	0	0	0	7	10	1	0	0
Q5 Prescribing	0	0	0	7	8	3	0	0
Q6 Medical record keeping	0	0	6	6	2	4	0	0
Q7 Recognising and working within limitations	0	1	3	7	5	2	0	0
Q8 Keeping knowledge and skills up to date	0	0	2	7	5	4	0	0
Q9 Reviewing / reflecting on own performance	0	0	4	4	4	6	0	0
Q10 Teaching (students, trainees, others)	0	0	3	6	2	7	0	0
Q11 Supervising colleagues	0	0	4	7	2	5	0	0
Q12 Commitment to care / wellbeing of patients	0	0	7	7	2	2	0	0
Q13 Communication with patients and relatives	0	3	9	3	0	3	0	0
Q14 Working effectively with colleagues	0	0	8	7	2	1	0	0
Q15 Effective time management	0	0	3	7	7	1	0	0

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree	Don't know	Blank (missing)	Spoilt
Q16 Doctor respects patient confidentiality	0	0	0	5	12	1	0	0
Q17 Doctor is honest and trustworthy	0	0	0	2	16	0	0	0
Q18 Performance not impaired by ill health	0	0	0	5	12	0	1	0

	Yes	No	Don't know	Spoilt	Blank (missing)
Q19 I am confident that this doctor is fit to practise medicine	16	0	2	0	0



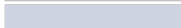
Blank, 'don't know' and spoilt responses are not included in the score calculation as presented in table 2.3

- insufficient responses to display data

## Evaluation question ratings and scores

Table 2.3: Mean percentage scores and benchmarks (Q1 - Q18)

	Your mean score (%)	Benchmark data (%) *				
		Min	Lower quartile	Median	Upper quartile	Max
Q1 Clinical knowledge	93	64	91	95	98	100
Q2 Diagnosis	93	58	89	93	96	100
Q3 Clinical decision making	90	55	88	93	96	100
Q4 Treatment including practical procedures	90	58	88	92	95	100
Q5 Prescribing	88	63	86	92	95	100
Q6 Medical record keeping	68	50	85	90	94	100
Q7 Recognising and working within limitations	75	50	87	91	95	100
Q8 Keeping knowledge and skills up to date	80	61	89	93	96	100
Q9 Reviewing / reflecting on own performance	75	55	85	90	93	100
Q10 Teaching (students, trainees, others)	73	50	84	91	95	100
Q11 Supervising colleagues	71	50	83	88	93	100
Q12 Commitment to care / wellbeing of patients	67	75	93	96	98	100
Q13 Communication with patients and relatives	50	59	88	93	97	100
Q14 Working effectively with colleagues	66	35	85	91	96	100
Q15 Effective time management**	81	48	80	87	91	100
Q16 Doctor respects patient confidentiality	93	69	94	96	98	100
Q17 Doctor is honest and trustworthy	97	75	94	97	99	100
Q18 Performance not impaired by ill health	93	50	93	96	98	100

	Your mean score for this question falls in the highest 25% of all means
	Your mean score for this question falls in the middle 50% of all means
	Your mean score for this question falls in the lowest 25% of all means

\*\* Effective Time Management (Q15) added to colleague questionnaire for Phase II data collection only.

- insufficient number of responses to generate score

See score explanation for percentage score calculation and quartile information

### \* Important notes about the benchmark data

The above benchmarks are based on all data that is available for the doctors who participated in our pilot work up to 2 July 2009. To calculate the benchmarks for each questionnaire item, we have included data only for doctors who had at least six valid responses returned for that particular item. We will update these benchmarks at regular intervals as the number of doctors undertaking these surveys increases.

Please bear in mind the following points when interpreting the above table:

**Purpose:** These benchmarks are provided to give you a sense of how you are performing in relation to other doctors who have completed the GMC surveys. They are not intended to imply any 'minimum standard' that doctors are expected to achieve for the purposes of revalidation.

**Sample size:** Of the doctors who have agreed to take part in the pilot work so far, 935 have completed a patient survey and 942 have completed a colleague survey. The benchmark data contained in this report is derived only from the doctors who returned sufficient numbers of patient or colleague responses. To date, 560 doctors have returned sufficient patient feedback and 620 doctors have returned sufficient colleague feedback.

**Voluntary participation:** It is not mandatory for any doctor to undertake these surveys. Therefore the benchmarks are based on a volunteer sample of doctors. As such, the benchmarks may be higher than might be expected if all doctors had contributed data.

**Range of practice:** The doctors who have contributed to the benchmark data work or provide care in a variety of settings and specialties. The above benchmarks relate to the whole sample of doctors who have contributed data, irrespective of their setting or speciality.

## Colleague comments

From the free text component of the questionnaire

### Comments

This doctor can be quite variable in his communication skills and manner with patients - excellent with some patients but poor with others.

I feel I am unable to answer some questions as I am not a qualified person.

It is difficult to complete this form as I do not work with this doctor - he has referred me a few patients who he has diagnosed appropriately. The referrals I receive from this doctor are usually appropriate.

This doctor's clinical knowledge and skills seem to be very good. However, his manner with staff and patients can sometimes be unfortunate.

I think the doctor is a hardworking and conscientious colleague.

Sometimes the doctor can come across a little abrupt or offhand. I don't feel it is intentional though - possibly he just doesn't react well to time pressures and stressful situations.

I am unsure how much reflective practice and keeping up-to-date this doctor is involved with, though he always attends and contributes to practice meetings and training sessions.

I have a good working relationship with the doctor and have never had any reason to seriously doubt their competence or professional abilities. Very experienced, dependable and trustworthy.

His note-keeping is sometimes not very detailed and health promotion/QOF boxes can be overlooked.

Generally a competent doctor who gets on OK with the rest of the team, but there are days when he is stressed and this does show.

# Self Assessment

## Comparison of self assessed scores with patient and colleague scores

**Table 3.1:**

Scores provided on a 1 - 5 scale where 1=Poor, 2=Less than satisfactory, 3=Satisfactory, 4=Good and 5=Very good

Patient questions	Your assessment	Patient assessment *
Q3a Being polite to patients	4	3.4
Q3b Making patients feel at ease in your presence	4	3.3
Q3c Listening to patients	4	3.5
Q3d Assessing patients' medical conditions	4	4.8
Q3e Explaining patients' conditions and treatment	4	4.6
Q3f Involving patients in decisions about treatment	4	4.1
Q3g Providing or arranging treatment for patients	4	4.7

Colleague questions	Your assessment	Colleague assessment *
Q1a Clinical knowledge	4	4.7
Q1b Diagnosis	4	4.7
Q1c Clinical decision making	4	4.6
Q1d Treatment (including practical procedures)	3	4.6
Q1e Prescribing	4	4.5
Q1f Medical record keeping	3	3.7
Q1g Recognising and working within limitations	3	4.0
Q1h Keeping knowledge and skills up to date	3	4.2
Q1i Reviewing and reflecting on own performance	3	4.0
Q1j Teaching (students, trainees, others)	4	3.9
Q1k Supervising colleagues	4	3.8
Q1l Commitments to care and wellbeing of patients	4	3.7
Q1m Communication with patients and relatives	4	3.0
Q1n Working effectively with colleagues	4	3.6
Q1o Effective time management	3	4.2
Q4a I respect patient confidentiality	5	4.7
Q4b I am honest and trustworthy	5	4.9
Q4c My performance is not impaired by ill health	4	4.7

- insufficient number of responses to generate score

-- no self assessment response provided

\* The Colleague/Patient assessment data represents the average score for the item, using all valid data provided by your patients / colleagues. Please note that your raw scores were transformed using the approach outlined in the section 'Supporting Documents' to obtain the mean percentage score referred to earlier in their report.

## Personal comment

From the free text component of the questionnaire

No personal comment provided




## Setting-specific benchmarks

For more information on how setting-specific benchmarks have been derived, please see the 'Background Information' document which is available on the appraiser training website.

## Patient setting-specific benchmarks - Primary Care

Table 4.1: Mean percentage scores and benchmarks (Q4,Q5)

	Your mean score (%)	Benchmark data (%) *				
		Min	Lower quartile	Median	Upper quartile	Max
Q4a Being polite	60	79	96	98	99	100
Q4b Making you feel at ease	59	79	95	97	98	100
Q4c Listening to you	62	75	95	97	98	100
Q4d Assessing your medical condition	94	75	93	96	98	100
Q4e Explaining your condition and treatment	90	81	93	95	97	100
Q4f Involving you in decisions	79	74	93	95	97	100
Q4g Providing or arranging treatment for you	94	77	94	96	98	100
Q5a Confidentiality of information	91	80	91	93	95	99
Q5b Doctor is honest and trustworthy	93	80	92	94	96	100

	Your mean score for this question falls in the highest 25% of all means
	Your mean score for this question falls in the middle 50% of all means
	Your mean score for this question falls in the lowest 25% of all means

- insufficient number of responses to generate score

See score explanation for percentage score calculation and quartile information

### \* Important notes about the benchmark data

The above benchmarks are based on all data that is available for the doctors who participated in our pilot work up to 2 July 2009 and work in your setting. To calculate the benchmarks for each questionnaire item, we have included data only for doctors who had at least six valid responses returned for that particular item. We will update these benchmarks at regular intervals as the number of doctors undertaking these surveys in your setting increases.

Please bear in mind the following points when interpreting the above table:

**Purpose:** These benchmarks are provided to give you a sense of how you are performing in relation to other doctors who have completed the GMC surveys and work in your setting. They are not intended to imply any 'minimum standard' that doctors are expected to achieve for the purposes of revalidation.

**Voluntary participation:** It is not mandatory for any doctor to undertake these surveys. Therefore the benchmarks are based on a volunteer sample of doctors. As such, the benchmarks may be higher than might be expected if all doctors had contributed data.




**Range of practice:** The doctors who have contributed to the above benchmark data work or provide care in a variety of settings and specialties. The above benchmarks relate to the whole sample of doctors who have contributed data within Primary Care.



## Colleague setting-specific benchmarks - Primary Care

Table 4.2: Mean percentage scores and benchmarks (Q1 - Q18)

	Your mean score (%)	Benchmark data (%) *				
		Min	Lower quartile	Median	Upper quartile	Max
Q1 Clinical knowledge	93	79	91	94	97	100
Q2 Diagnosis	93	77	88	92	95	100
Q3 Clinical decision making	90	66	88	93	96	100
Q4 Treatment including practical procedures	90	68	88	91	94	100
Q5 Prescribing	88	75	87	92	94	100
Q6 Medical record keeping	68	50	84	90	95	100
Q7 Recognising and working within limitations	75	50	87	90	94	100
Q8 Keeping knowledge and skills up to date	80	73	89	93	96	100
Q9 Reviewing / reflecting on own performance	75	63	86	90	93	100
Q10 Teaching (students, trainees, others)	73	63	84	90	94	100
Q11 Supervising colleagues	71	60	82	88	92	100
Q12 Commitment to care / wellbeing of patients	67	77	92	96	99	100
Q13 Communication with patients and relatives	50	69	88	93	97	100
Q14 Working effectively with colleagues	66	60	85	91	95	100
Q15 Effective time management**	81	48	81	87	91	100
Q16 Doctor respects patient confidentiality	93	84	94	97	98	100
Q17 Doctor is honest and trustworthy	97	81	95	97	99	100
Q18 Performance not impaired by ill health	93	50	94	96	98	100

	Your mean score for this question falls in the highest 25% of all means
	Your mean score for this question falls in the middle 50% of all means
	Your mean score for this question falls in the lowest 25% of all means

\*\* Effective Time Management (Q15) added to colleague questionnaire for Phase II data collection only.

- insufficient number of responses to generate score

See score explanation for percentage score calculation and quartile information

### \* Important notes about the benchmark data

The above benchmarks are based on all data that is available for the doctors who participated in our pilot work up to 2 July 2009 and work in your setting. To calculate the benchmarks for each questionnaire item, we have included data only for doctors who had at least six valid responses returned for that particular item. We will update these benchmarks at regular intervals as the number of doctors undertaking these surveys in your setting increases.

Please bear in mind the following points when interpreting the above table:

**Purpose:** These benchmarks are provided to give you a sense of how you are performing in relation to other doctors who have completed the GMC surveys and work in your setting. They are not intended to imply any 'minimum standard' that doctors are expected to achieve for the purposes of revalidation.

**Voluntary participation:** It is not mandatory for any doctor to undertake these surveys. Therefore the benchmarks are based on a volunteer sample of doctors. As such, the benchmarks may be higher than might be expected if all doctors had contributed data.

**Range of practice:** The doctors who have contributed to the above benchmark data work or provide care in a variety of settings and specialties. The above benchmarks relate to the whole sample of doctors who have contributed data within Primary Care

**Online File B: Appraisers' evaluations of feedback reports A to H – distribution and mean ratings (for overall assessment of report; concerns about doctor's performance; and acceptability of doctor's performance)**

	Report A N=100	Report B N=95	Report C N=93	Report D N=91	Report E N=88	Report F N=89	Report G N=87	Report H N=87
<b>Overall assessment of report</b>								
Excellent (5)	10 (10%)	0 (0%)	0 (0%)	0 (0%)	30 (34%)	1 (1%)	0 (0%)	34 (39%)
Very good (4)	<b>64 (64%)</b>	14 (15%)	1 (1%)	1 (1%)	<b>54 (61%)</b>	20 (22%)	5 (6%)	<b>48 (55%)</b>
Satisfactory (3)	23 (23%)	<b>42 (44%)</b>	<b>45 (48%)</b>	30 (33%)	4 (5%)	<b>58 (65%)</b>	<b>58 (67%)</b>	4 (5%)
Borderline (2)	3 (3%)	34 (36%)	38 (41%)	<b>44 (48%)</b>	0 (0%)	9 (10%)	22 (25%)	1 (1%)
Unsatisfactory (1)	0 (0%)	5 (5%)	9 (10%)	16 (18%)	0 (0%)	1 (1%)	2 (2%)	0 (0%)
<i>Mean (SD) rating</i>	3.8 (0.65)	2.7 (0.79)	2.4 (0.68)	2.2 (0.72)	4.3 (0.55)	3.1 (0.64)	2.8 (0.59)	4.3 (0.62)
<b>Concerns about doctor's performance</b>								
Not at all concerned (1)	<b>58 (58%)</b>	9 (9%)	0 (0%)	0 (0%)	<b>76 (86%)</b>	17 (19%)	5 (6%)	<b>64 (74%)</b>
Minor concerns only (2)	40 (40%)	<b>44 (46%)</b>	43 (46%)	27 (30%)	12 (14%)	<b>62 (70%)</b>	<b>55 (63%)</b>	23 (26%)
Significant concerns (3)	2 (2%)	42 (44%)	<b>48 (52%)</b>	<b>62 (68%)</b>	0 (0%)	10 (11%)	27 (31%)	0 (0%)
Extremely concerned (4)	0 (0%)	0 (0%)	2 (2%)	2 (2%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
<i>Mean (SD) rating</i>	1.4 (0.54)	2.3 (0.65)	2.6 (0.54)	2.7 (0.50)	1.1 (0.34)	1.9 (0.55)	2.2 (0.55)	1.3 (0.44)
<b>Acceptability of doctor's performance</b>								
Clearly acceptable (4)	<b>73 (73%)</b>	22 (23%)	7 (8%)	3 (3%)	<b>86 (98%)</b>	39 (44%)	15 (17%)	<b>80 (92%)</b>
Probably acceptable (3)	25 (25%)	<b>58 (61%)</b>	<b>63 (68%)</b>	<b>57 (63%)</b>	2 (2%)	<b>47 (53%)</b>	<b>63 (72%)</b>	7 (8%)
Probably unacceptable (2)	2 (2%)	15 (16%)	20 (22%)	27 (30%)	0 (0%)	3 (3%)	9 (10%)	0 (0%)
Clearly unacceptable (1)	0 (0%)	0 (0%)	3 (3%)	4 (4%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
<i>Mean (SD) rating</i>	3.7 (0.50)	3.1 (0.62)	2.8 (0.61)	2.6 (0.62)	4.0 (0.15)	3.4 (0.56)	3.1 (0.52)	3.9 (0.27)

Note: Emboldened figures represent modal response(s). SD = standard deviation.