The BJGP has for many years operated on an open peer review system, in which a minimum of two peer reviewers report on each original research article considered for publication and where the identities of the authors and reviewers are known to each other. Although peer review remains the ‘gatekeeper’ to research publication, its efficacy and reliability are still a topic of controversy. There is concern about the variation in the quality of peer review, both within and between journals. Editorial decisions such as the choice of reviewers, the interpretation of their comments, and the need to navigate between reviews offering divergent advice add to the difficulties. Formal training for reviewers is rare. Recently the ability of the system to identify fraud and plagiarism has been questioned.

A subjective scoring system was tested in a prospective observational study of the quality of 934 reviews.11 It consists of seven differently weighted subcategories reflecting both review content and format that contribute to the overall score, with a maximum of 14 points.8 Its validity was tested in a prospective observational study of 53 JVIR reviews. IRR was good (ICC = 0.84, P<0.001) but content validation was only assessed subjectively. There are several problems regarding the transferability of this system. Firstly, the JVIR’s reviewers use grade sheets to help evaluate the manuscript, which were incorporated into the instrument. Secondly, the instrument was tested predominantly by reviewers themselves, not journal editors. Lastly, the instrument could be criticised for the weighting of some of its attributes. For example, timeliness could contribute up to 3 points (21% of the total score). A mediocre review completed in a timely manner may therefore score relatively highly. Indeed, when this instrument was utilised by another group in a prospective observational study, the scoring was changed to reduce the relative importance of timeliness in the overall score.9

Finally, the Review Quality Instrument (RQI) was developed to be a simple, reliable, and valid scale for future studies of peer review by an editorial group from the BMJ.10 The scale rates seven aspects of the review, each on a 5-point Likert scale. The RQI underwent several revisions before being accepted, with testing of the tool occurring at each stage. It was subsequently used by the authors in a large randomised trial studying the effect of blinding and unmasking of reviewers on the quality of 934 reviews.13 From this it was shown that scores had a normal distribution, with no evidence of floor or ceiling effects. IRR of the total mean score was good (weighted k = 0.83), but was variable across subcategories (k = 0.49–0.73).

The worst-performing sections included scores for importance, constructiveness, substantiation, and interpretation. The authors suggested that better training and guidelines might improve the variability in raters’ scores. A further limitation is that this scoring system also puts a greater burden on editors’ time (2–10 minutes) than simpler scales. Despite these drawbacks, the RQI has been applied widely in randomised trials and observational studies.

IMPACT OF FEEDBACK ON REVIEW QUALITY

Only one article was identified that specifically addressed how the provision of written feedback alters subsequent review quality.12 This randomised trial carried out at the Annals of Emergency Medicine focused on
poor to moderate rated reviewers based on editors’ subjective quality rating (1–5 scale). Two feedback interventions were tested. The low-quality reviewers [median quality score of ≤3] received the standard journal measures, a brief summary of the specific content goals for a quality review, and the editor’s numerical rating of their review. The moderately rated reviewers [median quality score of 3 ≤4] experienced a more in-depth intervention, receiving in addition the editor’s ratings of the other reviews of the same manuscript and a copy of an exemplary review of a further manuscript. Neither intervention improved reviewer performance on subsequent manuscripts.

If anything, there was a more negative trend for the poorly rated reviewers, though this was not statistically significant. The authors concluded that the written feedback was ineffective. However, it is important to note that neither feedback intervention provided specific details of the problems with the review, and the subjective quality score is unlikely to have been meaningful for the reviewers. Additionally, the reviewers involved in the study did not self-select; it is likely that feedback is a more effective educational tool for those who actively seek it.

**SURVEY OF B JGP REVIEWERS**

Given the large body of literature in educational research to suggest that feedback does improve performance and the current culture of self-improvement within medicine, we still felt this was a timely opportunity to use feedback in the improvement of peer review at the individual level. We surveyed *B JGP* reviewers to determine where the focus of feedback should lie, the form it should take, and who would most benefit from it.

We invited 120 *B JGP* reviewers with a range of experience to complete a short online questionnaire regarding feedback for peer review. There was a 58% response (70 reviewers). Although reviewers with different levels of experience responded, including those who had not yet reviewed for the journal, there was little difference in the responses between those who were more and those who were less experienced.

Most reviewers (93%) said that they would value feedback for their reviews and would find it useful for improving future reviews. A further 65% felt that feedback on every review would be the most appropriate, although there was a significant subgroup (17%) who said that only initial reviews warranted feedback. Individualised written feedback was the most popular format (47%), followed by a number of scores in subcategories (31%). A single score was the least popular (21%).

The questionnaire also allowed free text responses, with several common themes emerging. Reviewers felt the current system of receiving decision letters and a copy of other review(s) was already very helpful. While routine personalised feedback would be greatly appreciated and useful for professional development, there was recognition that this may overload the journal. Feedback for novice reviewers and annual reports were among suggestions of valuable forms of feedback. There was also support for peer review workshops as a training tool.

**A NEW FEEDBACK SYSTEM FOR THE B JGP**

Taking into account the literature review, the questionnaire results, the *B JGP*’s AllenTrack submission software, and the time and work burden of providing feedback by the editorial staff, the *B JGP* will now be introducing a new feedback system which will offer:

- routine provision of feedback on a 5-point scoring system based on the system devised by the editors of *Obstetrics and Gynaecology* (Box 1), adapted for the A, B, C, D, E system already built into AllenTrack, which allows reviewers to see their mean review rating; and
- narrative feedback on request on up to two occasions per year, providing comments along the lines of the review criteria, expanding and advising where necessary. This may be of particular interest to new reviewers.

**CONCLUSION**

Peer review is an imperfect system, but is probably the best method we have to safeguard original research publication. We hope that the new feedback system implemented at the *B JGP* will go some way to improving the quality and consistency of the journal’s own peer review process.

Abigail Moore,
Foundation Doctor, Oxford University Hospitals, Oxford, UK.

Roger Jones,
*B JGP* Editor, RCOP, London.

Provenance
Freely submitted; not externally peer reviewed.

©British Journal of General Practice
This is the full-length article [published online 30 Jun 2014] of an abridged version published in print. Cite this article as: *Br J Gen Pract* 2014; doi: 10.3399/bjgp14X680713.
REFERENCES