# What does better peer review look like? Findings from 132 journals completing a self-assessment exercise

Thomas E. Gaston[1], Chris Graf[1], Elisha Morris[1], Elizabeth Moylan[1], Sarah Pedder[1], and Michael Willis[1]

[1]John Wiley & Sons

April 28, 2020

**Abstract**

*Aim:* We wanted to understand how well journal teams, comprising editors, managing editors, reviewers and publishers, perform across five Essential Areas of peer review according to a self-assessment of their own editorial and peer review processes. We also wanted to identify and share the best practices that journals use and recognise potential obstacles that could be overcome.

*Methods:* Journals used a Self-Assessment tool to assess their peer review processes by answering questions and giving themselves a quantitative score and providing a qualitative explanation for their rating, across the five 'Essential Areas' of Integrity, Ethics, Fairness, Usefulness and Timeliness. Wiley colleagues independently rated the journals to distinguish best practices and identify potential obstacles.

*Results:* We examined the responses of 132 journals which completed the Self-Assessment exercise. Journals tended to rate themselves more highly than the study authors did. The greatest variation in rating between journal self-rating (SA-score) and the study authors' rating (R-score) was in the Essential Area of Usefulness, with the smallest variation in the area of Ethics. We identified a set of best practices that could help improve peer review in each of the Essential Areas.

*Conclusion:* The Self-Assessment encourages journals to reflect on and change their peer review processes and offers practical guidance on how to do this. They benefit from greater awareness of technical solutions that exist to help them in this. The Self-Assessment also highlights how journals can be inconsistent in the way that their processes operate, with one policy in place for authors and a different or no policy in place for reviewers/editors. Rather than be content with the status quo, journals should strive to improve processes in the light of changing community expectations and technological advances.

Correspondence: Michael Willis, John Wiley & Sons, 9600 Garsington Road, Oxford, OX4 2DQ, United Kingdom. Email: miwillis@wiley.com.

## Key points

- Journals can assess their performance in the peer review process using a self-assessment tool which encourages them to aspire to continuous improvement of their processes.
- Journals tend to assess their own performance more positively than is warranted.
- Journals should focus on ensuring transparency and consistency in the guidelines and processes they provide to authors, editors and reviewers.

# Introduction

Every innovation starts with a key problem or question to solve. Our starting point was a question from a Wiley colleague: What does gold standard peer review look like? This led to other questions: How can a journal team (comprising academic editors, managing editors and publishers) really know whether the peer review 'service' they deliver to researchers is as good as that delivered by others? How can they identify their strengths and address their weaknesses? How can they differentiate in ways that really matter to researchers?

To answer these questions a team of colleagues began by reviewing literature on the topic and collecting and analysing 40 case studies in peer review. The cases were submitted by a range of managing editors, academic editors and publishers employed by or working with Wiley spanning different disciplinary areas and geographical regions. The team identified hallmarks of better peer review and defined five 'Essential Areas' – integrity, ethics, fairness, usefulness and timeliness (Table 1).

| Essential Area | Definition |
|---|---|
| **Integrity** | Peer review addresses the integrity of the work under review when it focuses on ensuring that researcher |
| **Ethics** | Peer review addresses the ethics of the work under review when it establishes that the work was conduct |
| **Fairness** | Peer review is conducted fairly when it considers papers on their own merit, without regard for the ident |
| **Usefulness** | Peer review is useful when it benefits all stakeholders in the process. It means providing constructive fee |
| **Timeliness** | Peer review is conducted in a timely manner when an outcome is reached quickly, without compromising |

**Table 1.** Definitions of the five Essential Areas of peer review.

The team described best practice in each Essential Area in detail, using the case studies and findings from the literature review, and developed a checklist of questions to help journal teams. The work was shared as a preprint in April 2018 (Allen *et al* ., 2018) and published after peer review and revision (Allen *et al.,* 2019).

At the same time a smaller group of colleagues created an online tool for journals to use, the 'Better Peer Review Self-Assessment, Version 1.0', developed from the checklist presented in Appendix 1 of the preprint. The Self-Assessment was built on Microsoft FormsTM and Microsoft FlowTM, with a dashboard built using Microsoft ExcelTM that partially automates the creation of further feedback in the form of a Better Peer Review Self-Assessment Quartile, Badge, and Data Visualization. Sixteen colleagues attended three workshops in September 2018 in Wiley offices in the USA (Hoboken) and UK (Chichester and Oxford). They completed the Self-Assessment and shared their feedback, which led to improvements to the Self-Assessment. This was then made available to Wiley colleagues as Version 2.0 and is now accessible to everyone, including individuals outside Wiley (https://wiley.com/go/betterpeerreview).

The Better Peer Review Self-Assessment comprises three steps. First, in the 'Think and Reflect' step, journals answer 48 questions focused on the five Essential Areas. Journals decide how to approach the Self-Assessment: individual team members can look at the questions prior to undertaking the Self-Assessment as a group or, if preferred, an individual journal team member such as a Managing Editor can first complete the Self-Assessment, then discuss the results with the rest of the team, revise practices accordingly, and then repeat. Journals must answer the question and also briefly explain the rationale for their answer with a free

text summary. Next, in the 'Immediate Feedback' step, the journal receives instant on-screen feedback as well as an immediate record by email of their answers. Finally, in the 'Summary Feedback' step, journals receive another follow-up email with detailed information on their Quartile compared with other journals that have completed the Self-Assessment, Badge, Data Visualization, and some hints and tips for the journals to use, if they wish, to improve their processes. The 'Badge' is a radar plot illustrating performance (Figure 1). The Data Visualization breakdown is a histogram comparing a journal's scores with the mean scores reported by all other journals (Figure 2). The Better Peer Review Self-Assessment therefore enables journal teams to identify their strengths and weaknesses; to find out how their practices across the Essential Areas compare with those of their peers; and to receive guidance about how they might improve their processes.


Figure 1: An example of the Better Peer Review Badge awarded to journals after they have completed the self-assessment.

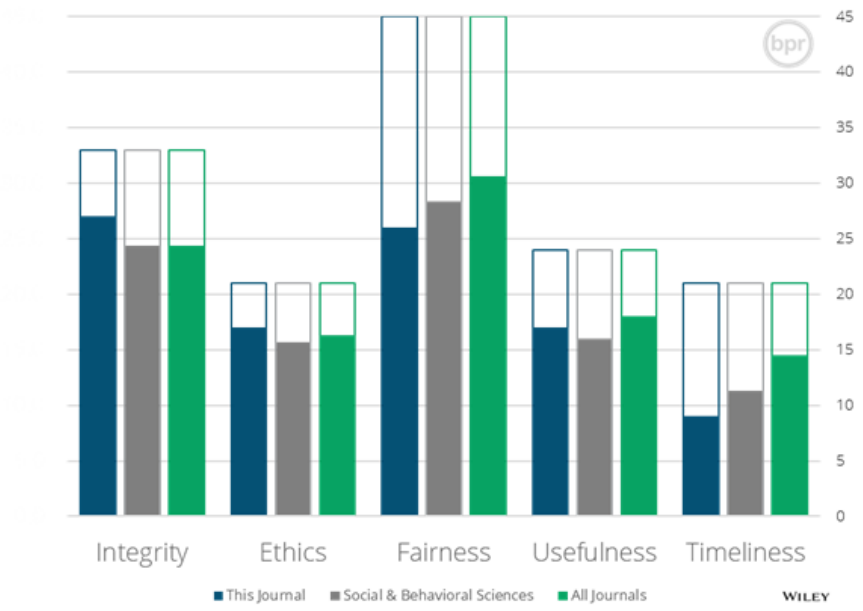


Figure 2: An example summary of the Better Peer Review data visualization that a journal receives after completing the self-assessment

We wanted to learn how the initiative was working, to identify and report on best practices that journals use, and to identify potential obstacles that might be overcome.

# Methods

Within the Better Peer Review Self-Assessment tool, journal teams were asked to rate their current practice with a self-assessment score (SA-score) from one to three (Table 2). Additionally, journal teams were asked to reflect on their response and and briefly explain a free text summary why they gave themselves each score. This provided respondents with an opportunity to describe specific journal practices. The subject areas and ownership represented by the 132 journals that completed the self-assessment are described in Supplementary Table 1.

| **Self-Assessment Score (SA-score)** | **Interpretation** |
|---|---|
| 3 | If your answer is "yes" |
| 2 | If you think you "could do better" |
| 1 | If your answer is "no" |

**Table 2:** Rating system that individuals in journal teams used to assign their own self-assessment scores (SA-score)

In order to synthesise and describe processes that define best practice in peer review and to define potential obstacles encountered by journal teams, five raters (TG, CG, ECM, EM, MW) also rated the answers to each question using the qualitative feedback provided by the journal teams. The journal responses were

anonymised with a journal number assigned to each journal. The questions were then divided between the raters according to the five Essential Areas, with one rater assessing and rating every journal response within a specific Essential Area. The raters followed a consistent rating system (Table 3) devised after the questions and after the self-assessments had been completed. The raters scored the responses from one to three, and in addition rated a non-applicable answer as 'N/A' and an unaccountable answer as 'U'. The raters' scores are referred to as Rated scores (R-scores).

| **Rated Score (R-score)** | **Interpretation** |
|---|---|
| 3 | Better practice: answer indicates that what is done goes beyond the bare minimum, indicating |
| 2 | Bare minimum practice: answer only indicates what is covered by generic publisher policies, by |
| 1 | Practice not done/done insufficiently: answer indicates that this is not currently done or is not |
| U | Unaccountable: answer is unintelligible, or unrelated to the question, or insufficient justification |
| N/A | Not applicable: answer indicates the question is irrelevant to the journal, for example because i |

**Table 3:** Rating system used by Wiley colleagues for analysis (R-score)

To eliminate bias in the rating process, journal identifiers, including subject areas, were not revealed to any of the raters until after the initial qualitative and quantitative analysis had been completed.

To improve inter-rater reliability, each rater flagged answers that needed further discussion with the other raters. In addition to rating answers, the raters also highlighted examples of interesting and exceptional practice that would form the basis of identifying quality peer review. They also highlighted examples of potential obstacles preventing improvements in a given area.

Once all journal answers had been rated, another team member (SP) who had not been involved in the rating process carried out further qualitative and quantitative analysis.

The SA-score for each journal's response was subtracted from the R-score, the difference enabling us to assess journals' levels of understanding or awareness, which in turn could help us evaluate how the Self-Assessment is working (Supplementary Table 2). The scores and the differences were assessed by journal subject area, by Essential Area of peer review, and by each question within the Essential Area.

We analysed the qualitative responses to determine best practice and obstacles to good practice. To find examples of better peer review, we extracted the highlighted responses with an R-score of three, and to find obstacles to better peer review, we extracted the highlighted responses with an R-score of one. We also extracted answers scored 'N/A' to determine how questions might be applicable only to certain subject areas. From this we produced a synthesised set of best practice recommendations for each Essential Area, as well potential obstacles to good practice in peer review. We published this online at https://secure.wiley.com/better-peer-review, with a interactive infographic to help editors and researchers explore ways in which they can foster and experience better peer review.

# Results

## Quantitative analysis

132 journals across a range of disciplinary areas completed the Self-Assessment, resulting in a total of 6,336 responses for the 48 questions. Each journal took an average of 69 minutes to answer the 50 questions in

the Self-Assessment. The subject areas represented by the journals are shown in Table 4.

| Subject area | Number of journals |
|---|---|
| Health Sciences | 51 |
| Life Sciences | 34 |
| Physical Sciences & Engineering | 14 |
| Social Sciences & Humanities | 33 |
| **Total** | **132** |

**Table 4:** Journals by subject category.

The five raters identified 385 free text responses that were rated 'unaccountable' and 140 that were rated 'not applicable'. These responses were excluded from the subsequent qualitative analysis.

94 journals recorded a mean SA-score of 2 or higher, whereas we recorded 45 journals with a mean R-score of 2 or higher. Mean SA-scores overall were higher than R-scores (2.17 vs. 1.87; difference = -0.30). Scores varied only marginally by subject category; in every case SA-scores were higher than R-scores, with R-scores less variable across subject area (SA-scores range = 0.23, compared with R-scores range = 0.14; Supplementary Table 3).

We found greater variation when considering the scores by Essential Area (Table 5). The highest average SA-score was for Ethics, with R-scores differing only marginally, whereas rhere was a much larger difference between the SA-scores and R-scores for Usefulness.

| Essential Area | Essential Area | Mean SA-Score | Mean R-Score | Difference |
|---|---|---|---|---|
| Ethics | 2.33 | 2.30 | 0.03 | 0.03 |
| Integrity | 2.22 | 1.93 | 0.28 | 0.28 |
| Fairness | Fairness | 2.05 | 1.74 | 0.31 |
| Timeliness | Timeliness | 2.10 | 1.71 | 0.40 |
| Usefulness | Usefulness | 2.28 | 1.83 | 0.44 |
| **Overall** | **Overall** | **2.17** | **1.87** | **0.30** |

**Table 5:** Mean SA- and R-scores by Essential Area, ranked by difference.

In assessing the highest mean R-scores across all Essential Areas (Table 6), the Essential Areas of Usefulness and Timeliness did not feature in the top five. The highest scored Usefulness question was Question 32, 'Do you seek feedback from editors about the editorial process, and act on that feedback?' (10th; mean R-score = 2.21). In the Essential Area of Timeliness, the highest scoring question was Q35, 'Do you conduct regular reviews of your workflow and metrics, and do you act on the findings to make improvements?' (14th; mean R-score = 2.17). For the five questions that scored most highly, four have standard system solutions, i.e. they are enabled by default in the journals' submission systems.

| Question | Es |
|---|---|
| 27. Do you check for overlapping text and potential plagiarism? | Et |
| 43. Do you copy all authors on emails to acknowledge receipt of their manuscript? | Fa |
| 28. Do you have a plan for how to act when you identify a potential research integrity and publishing ethics issue? | Et |

| Question | Es |
|---|---|
| 38. Do you ask authors to declare conflicts of interest, and do you publish in their article a statement regarding these? | In |
| 37. Do you ask authors to declare sources of funding, and do you publish in their article a statement regarding these? | In |

**Table 6:** Highest scoring questions based on R-scores.

In assessing the lowest mean R-scores across all Essential Areas (Table 7) neither Ethics or Integrity featured in the lowest five. The lowest rated Ethics question was Q24, 'Do you check images for inappropriate manipulation?' (41st; mean R-score = 1.42). The lowest rated Integrity question was Q4, 'Do you explain to reviewers how their contribution is used to decide whether to publish the manuscript?' (40th; mean R-score = 1.46). Of the five questions with lowest R-scores, none has a system solution. For example, submission systems do not currently have an automated way to present key editorial metrics to authors (Q49, Table 7).

| Question |
|---|
| 49. Do you share with authors key editorial metrics such as median time to decisions and acceptance rates? |
| 13. Do you discourage reviewers from raising unreasonable additional issues at re-review? |
| 15. Do you attempt to manage the effects of implicit bias in peer review, related to the nationality, religious or political bel |
| 29. Do you collect and declare potential conflicts of interest for your editors on the journal website? |
| 19. Do you seek feedback from reviewers about the editorial process, and act on that feedback? |

**Table 7:** Lowest scoring questions based on R-scores.

The number of questions with R-scores of 'N/A' were highest among Social Sciences and Humanities journals (3.72%; $n = 59$) and lowest among Health Sciences journals (1.31%; $n = 32$) (Supplementary Table 4). The question with the highest number of 'N/A' responses related to image checking ($n = 50$), followed by questions referring to ethics statements and reporting guidelines (Table 8).

| Question |
|---|
| 24. Do you check images for inappropriate manipulation? |
| 39. Do you ask authors to state the ethical standards required for experiments performed on animals/humans, and do you p |
| 3. Do you refer reviewers to reporting guidelines such as CONSORT, PRISMA, etc. (http://www.equator-network.org/)? |
| 36. Do your author guidelines explain the minimum requirements for completeness of reporting their research (e.g. CONSO |

**Table 8:** Questions with a high number of non-applicable responses ('N/A').

In some instances, we found inconsistencies between how journals framed questions to authors and to peer reviewers. For example, only one question (Q3, on referring reviewers to reporting guidelines) received no R-scores of 3, whereas ten journals received an R-score of 3 for Q36 (on referring authors to reporting guidelines), and of these ten, four had an R-score of 1 for Q3.

In some instances, we found a discrepancy between how journals rated themselves and the practices they undertook. For example, one journal had a SA-score of 1 for Q27 on plagiarism detection software but has

iThenticate text similarity software incorporated into its submission system, perhaps indicating a lack of awareness of the technology or a misunderstanding of the question.

Of the 132 journals, 10 journals of the 49 operating double-blind peer review had a SA-score of one for question 15 about how they address bias in peer review. In contrast, 7 journals of the 83 operating single-blind peer review had a SA-score of three.

Calculating average Timeliness scores for all journals and dividing them in quartiles allowed comparisons to be made with actual journal times from submission to first and final decision. We found no correlation between average turnaround times and SA-scores (Table 9). We did, however, find a correlation between average turnaround times and R-scores for Timeliness, with the shortest turnaround times correlating to the highest scores for Timeliness (Table 10).

| **Quartile** | **Mean time to first decision (calendar days)** | **Mean time to final decision (calendar days)** |
|---|---|---|
| Q1 (lowest) | 62.21 | 126.28 |
| Q2 | 81.48 | 155.22 |
| Q3 | 61.60 | 105.74 |
| Q4 (highest) | 65.02 | 126.90 |

**Table 9:** Mean turnaround times by mean Timeliness SA-scores

| | **Mean time to first decision (calendar days)** | **Mean time to final decision (calendar days)** |
|---|---|---|
| Q1 (lowest) | 73.97 | 155.01 |
| Q2 | 72.56 | 129.46 |
| Q3 | 65.57 | 119.31 |
| Q4 (highest) | 58.61 | 118.02 |

**Table 10.** Mean turnaround times by mean Timeliness R-scores

Examples of best practice as identified by each rater across the five areas are shown in Table 11 and have also been published through our interactive online tool at https://secure.wiley.com/better-peer-review. Potential obstacles that were preventing improvements in a given area were also identified. These included a lack of technical solutions to a particular issue, a lack of clear policy and implementation and a lack of awareness of understanding why a particular change is necessary (Table 11). **Table 11:** Self-assessment questions with Essential Area, mean SA-scores and R-scores, guidance for best practice and potential obstacles to implementing best practice.

## Qualitative analysis

We report on the qualitative analysis by Essential Area.

*Integrity*

Best practice in the Essential Area of Integrity (Table 11) focused on having transparent policies for researchers with respect to data sharing and reporting guidelines, and on asking for this information on submission. Requiring researchers to share information on their funding sources and potential conflicts of interest, and providing these in the published article, also ensures that reviewers and readers form a complete assessment of the research presented. The quantitative analysis suggested that compliance with these recommendations is good. Most journals ask authors to declare funding (Q37: 46% R-score = 2; 43% R-score = 3) and conflicts of interest (Q40: 40% R-score = 2; 51% R-score = 3). Compliance on data policies is more mixed (Q22: 54% R-score = 2; 29% R-score = 1). Even excluding journals for which reporting guidelines are not applicable (14%), a significant number have no policy on using such guidelines to ensure minimum requirements for completeness.

Journals also need to direct reviewers to the sources of information they require to ensure that the research they are peer reviewing is completely and accurately reported. From the qualitative responses shared, suggestions for how to do this included providing specific questions on the reviewer report form that direct the reviewer to the methodology used, and links to relevant reporting guidelines where applicable. Compliance with these recommendations is low; 55% of journals do not explain to reviewers how their contribution will be used to make a decision (Q4:R-score = 1); 44% do not direct reviewers to assess the methodology (Q6: R-score = 1); and 46% do not refer reviewers to reporting guidelines, such as CONSORT (Q3: R-score = 1).

There is also a need to support reviewers by explaining how their contribution facilitates the editorial decision and how they can provide a constructive report, for example with appropriate information in the initial invitation email, reviewer report form and reviewer guidelines. Journals going the extra mile also provided guidance via editorials, newsletters and presentations. It was recognised that editors too need guidance on their approach to peer review, for example by using a variety of tools or approaches to ensure appropriate reviewers are invited to peer review and that confidentiality of the peer review process is maintained. It was recognised that editorial criteria for consistent decision making should be discussed regularly and that new editors could receive mentoring from more experienced editorial board members. 48% of journals had no formal criteria for editors to make consistent editorial decisions (Q23: R-score = 1).

If journals are committed to upholding the integrity of the research they publish, there needs to be a means by which readers can raise concerns. 58% of journals had some mechanism to achieve this (Q21: R-score = 2) and 28% had good practice on this (Q21: R-score = 3).

Some of the challenges apparent in these areas were a lack of coordinated approach, for example referring authors to reporting guidelines, but not peer reviewers (Moher, 2018). This was also apparent in requests for authors to provide funding sources or potential conflicts of interest, but not sharing the information in the published article. There was also a reluctance on the part of some journals to risk patronising reviewers about requirements for peer review, perhaps fearing resistance or accusations of 'hand-holding,' and an uncertainty about who is responsible for providing training, with opinions divided between the institution or publisher.

*Ethics*

Best practice in this Essential Area of ethics (Table 11) focussed on the need for transparency in relation to policies explained in the journal's author guidelines (for example in relation to preprints, confidentiality of the review process and any journal checks that were conducted). In terms of ensuring ethical standards of published research, journals with best practice asked authors to provide details of any ethical approvals needed for their research, together with ethics committee approval and reference numbers where applicable. Journals reaching high standards also used appropriate tools to check for potential plagiarism or image manipulation.

Ethics was the highest scoring Essential Area in our quantitative analysis, with 88% of journals operating good practice on checking for overlapping text (Q26: R-score = 3), 73% having a plan of how to respond to ethics issues (Q28: R-score = 3), and 47% have a preprint policy (Q26: R-score = 3). 41% of journals did not check for image manipulation even after excluding the 38% non-applicable journals (Q24: R-score = 1), and 33% journals not explaining their approach on peer review confidentiality to editors (Q25: R-score = 1).

Obstacles to improving peer review in this area included the lack of technology solutions (for example with respect to checking for image manipulation) or not setting clear policies (for example with respect to preprints).

*Fairness*

Good practice with respect to this Essential Area of fairness (Table 11) is again focussed on transparency around journal policies *and*having a procedure or process in place for implementing the policy. For example, on authorship, handling potential conflicts of interest (of editors and peer reviewers) and maintaining confidentiality of the peer review process.

While many journals ask reviewers to declare conflicts of interest (Q9: 56% R-score = 2), 75% do not collect editor conflicts of interest (Q29: R-score = 1) and 51% have no policy on how to handle such editor conflicts (Q30: R-score = 1).

64% of journals have a policy on authorship (Q41: R-score = 2) but 66% do not ask for a description of author contributions (Q42: R-score = 1).

64% of journals have some method of explaining peer review confidentiality to authors (Q40: R-score = 2) but 27% do not (Q40: R-score =1).

There was also a need to encourage fairness in the peer review process and to discourage peer reviewers from raising unreasonable additional concerns at re-review.

Some journals took a proactive approach to reviewer diversity and particularly encouraged this when peer reviewers were invited. Although journals recognise the benefits of a diverse editorial board in helping to foster diversity of peer reviewers (Ortuzar, 2019), the majority of journals (77%) have no practice on encouraging diversity when selecting reviewers (Q12: R-score = 1); only 12% have good practice (Q12: R-score = 3). 70% have no means for attempting to manage the effects of implicit bias (Q15: R-score = 1).

Obstacles to better practice included an assumption that there was already an awareness of the particular issue, for example with respect to the need to maintain confidentiality of the peer review process, implying there would only be a need to share information if asked. Other obstacles included a lack of understanding about why a change was necessary, for example with respect to declaring editors' potential conflicts of interest on journal websites; and the lack of clear policy or process, for example with respect to considering diversity when inviting peer reviewers.

*Usefulness*

Recognised best practice in the Essential Area of usefulness (Table 11) involves providing opportunities for further feedback on the review process, whether to authors, in the form of additional comments from editors on the decision, or to reviewers, in terms of feedback on the decision reached. Feedback is a two-way process, and journals with good practice should also encourage feedback on the peer review process from authors

and peer reviewers too (either in decision emails or in surveys). Providing additional information where it was most helpful to do so was also recognised as a useful practice, for example checklists for submission in author guidelines or an editorial on how to write a manuscript suitable for publication in the journal. Support for editors and training was mentioned, especially with regard to evaluating manuscripts and peer review. The importance of recognising the voluntary work reviewers do, whether through services such as Publons, certificates, awards or discounts on publisher products, was also apparent.

The main obstacles to best practice in this area were a lack of awareness that seeking feedback could be beneficial or a concern that this could further burden reviewers or even the journal. Many journals acted if feedback was received, although they did not actively seek it. Another general theme was an inconsistent approach to the sharing of information. Some but not all reviewers may, for example, be informed about an editorial decision, and some but not all reviewer contributions were recognised.

36% have good practice on seeking feedback from editors (Q32: R-score = 3), and 45% have some practice on this (Q32: R-score = 2). On the other hand, 74% have no practice around soliciting feedback from reviewers (Q19: R-score = 1).

*Timeliness*

Best practice in the Essential Area of timeliness (Table 11) focussed on transparent communication both on journal websites and in correspondence with authors about expectations regarding potential timelines, including average decision times and information on acceptance rates. Reviewers should have the ability to let the journal know if they are unable to assist or if their report will be delayed. While automatic reminders and automatic updates can be helpful, if a manuscript is unreasonably delayed, a personal email explaining the circumstances to the author is essential. It is helpful for journal teams to monitor journal turnaround times on a regular basis to make any immediate or long-term adjustments as necessary to workflows. Additional tools may accelerate elements of the submission and peer review process, such as tools to detect textual overlap or to find potential peer reviewers.

Obstacles with respect to best practice in this area include limitations in technology. For example, while it is always possible to capture information that a reviewer has declined to review, it may not always be possible to capture the reason why the reviewer declined. Other obstacles related to a lack of awareness that particular tools exist or prioritising timeliness given the lack of apparent targets or reporting on journal metrics, or an unwillingness to share information on journal metrics outside of journal teams.

68% of journals do not share key editorial metrics with authors (Q49: R-score = 1), 52% do not describe the stages used in peer review (Q48: R-score = 1), and 56% do not inform authors when they might experience delays (Q50: R-score = 1). On the other hand, 53% have some practice about sharing timeliness goals across the journal teams (Q34: R-score = 2) and 17% have good practice (Q34: R-score = 3); 44% conduct regular reviews (Q35: R-score = 2) and 23% have good practice (Q35: R-score = 3).

*General comments*

The qualitative responses show that journals can and should aspire to making improvements to their processes: no journal scored highly in every respect. To take just one example of opportunity for improvement from each of the Essential Areas:

- Integrity: 46% of journals do not refer reviewers to reporting guidelines
- Ethics: 41% of journals for whom it is relevant do not have a process in place for checking for potential image manipulation
- Fairness: 75% of journals do not collect editor conflicts of interest

- Usefulness: 74% of journals have no practice around soliciting feedback from reviewers
- Timeliness: 56% of journals do not inform authors when they might experience delays

It is encouraging to see that the highest ratings were in the area of Ethics, no doubt because this area has received the most attention and is arguably where the stakes are highest; journals which have weak ethics processes in place stand to lose a great deal if their poor practice is exposed.

It was instructive, but not surprising, to note that there were differences between journals' self-rating and our rating. Journals published by large publishing houses may not always be aware of systems and processes which the publisher has implemented on a wide scale. Publishers could invest more effort in helping journals stay aware of these developments.

Some aspects of the peer review processes are checked for in submission systems and so higher compliance is expected. For example, it is now standard practice across Wiley journals to use iThenticate to check for overlapping text (Q27). For certain questions it easier than others to perform well. Ethics questions score highly, perhaps because so much is at stake in having poor practice in this area, and it is a more regulated area with guidelines and recommendations for good practice.

From our analysis of the qualitative responses we identified a number of obstacles to good practice. These included a lack of technical knowledge or awareness of the opportunity to adopt good practice, for example in not being familiar with readily available technological solutions; a lack of consistency, for example asking authors to comply with reporting guidelines but not asking reviewers to assess manuscripts against these; a fear of additional workload, for example in not offering authors the opportunity to appeal against decisions; and a fear of exposure, for example in having weaknesses in the peer review process identified and called out.

Rather than it being an end in itself, we view the self-assessment as being the beginning of a journey. Journals can use the badges they receive as part of the exercise to identify strengths and areas for improvement, and then be guided by the hints and tips infographic (http://secure.wiley.com/better-peer-review) to make adjustments to their processes. Other suggestions for follow-up include discussion of some of these practices at editorial meetings, for example in exploring how to make editorial boards more diverse; providing informal or formal training for editors and reviewers; and repeating the self-assessment after, say, six months or one year, to assess progress.

*Study limitations*

With 132 journals providing responses the sample size was not large enough for us to make wide-ranging assumptions about journal practice. Also, not all questions in the self-assessment are applicable to all subject areas. A further limitation was the fact that we defined the criteria for our own rating of 1-3 which we then applied consistently across all journals in the study, whereas each journal rated itself independently of any other journal and with no specific criteria against which to judge. We also added 'u' and 'n/a' categories for our rating, which were not available to journals for the self-assessment. Journals were also operating independently of each other and not necessarily familiar with common practice and system solutions, whereas we had the benefit of observing what was a common or acceptable standard across multiple journals, or what practices were in place simply because of available system solutions.

# Conclusion

All journals, regardless of discipline, business model, publisher or location, benefit from completing the Self-Assessment tool which is freely available to the scholarly journal publishing community. It gives journals an opportunity to reflect on their peer review practices and to consider how they can improve in areas where they

may be relatively weak, in the spirit of providing greater quality in the practice of peer review. The model responses illustrating best practice, which we have synthesized can be shared with journals after they have completed the Self-Assessment, provide an invaluable resource for other journals to guide them in providing a high quality peer review service for their authors. Rather than being complacent about their current practice, and acknowledging that most journals may employ best practice in some areas, journals always have room for improvement, particularly with changing expectations in subject communities and advances in technology. In this respect there is a great opportunity for publishers to be influencers and early adopters of process change. The Self-Assessment has limitations, notably that not all questions are appropriate to all subject communities, however, we believe the general approach can help journal teams – whether editors, managing editors, reviewers or publishers – to reflect in depth on their peer review processes, to identify areas of weakness in those processes, to highlight gaps in knowledge of technical solutions that exist to improve the processes, and to draw attention to inconsistencies in the way that journals communicate with authors, reviewers and editors.

## Acknowledgements

Thank you to the Wiley team members who have made contributions since September 2018 when work on the Self-Assessment itself began, and to those who gave feedback at the workshops. Thank you to Wiley colleagues Thomas Broomfield, Tom Dent and Ron Perazza for their help in creating the self-assessment tool, and to Claire Mayne for marketing support. Lastly, we are most grateful for the 132 journal teams that completed the Self-Assessment so far, since we launched in December 2018.

## Author Contributions

Thomas Gaston: Conceptualization, formal analysis, methodology, validation, visualization, writing – original draft, writing – review and editing; Chris Graf: Conceptualization, formal analysis, methodology, supervision, validation, writing – original draft, writing – review and editing; Elisha Morris: Conceptualization, data curation, formal analysis, methodology, project administration, validation, visualization, writing – original draft, writing – review and editing; Elizabeth C Moylan: Conceptualization, formal analysis, methodology, validation, writing – original draft, writing – review and editing; Sarah Pedder: Conceptualization, formal analysis, methodology, validation, visualization, writing – original draft, writing – review and editing; Michael Willis: Conceptualization, formal analysis, methodology, validation, writing – original draft, writing – review and editing.

## References

Allen, Heidi, Emma Boxer, Alexandra Cury, Thomas Gaston, Chris Graf, Ben Hogan, Stephanie Loh, et al. 2018. "What Does Better Peer Review Look Like? Definitions, Essential Areas, and Recommendations for Better Practice." OSF Preprints. April 20. https://doi.org/10.31219/osf.io/4mfk2

Allen, H., Cury, A., Gaston, T., Graf, C., Wakely, H., Willis, M. (2019) What does better peer review look like? Underlying principles and recommendations for better practice. *Learned Publishing* 32 (2), 163–175. https://doi.org/10.1002/leap.1222

Graf, C. (2019, March 13). Ethics: An Essential Area for Better Peer Review. Retrieved from https://www.wiley.com/network/latest-content/ethics-an-essential-area-for-better-peer-review, 12 September 2019

Graf, C. (2019, March 20). Fairness: An Essential Area for Better Peer Review. Retrieved from https://www.wiley.com/network/latest-content/fairness-an-essential-area-for-better-peer-review, 12 September 2019

Graf, C. (2019, March 5). Integrity: An Essential Area for Better Peer Review. Retrieved from https://www.wiley.com/network/journaleditors/peer-review-and-editorial-office-management/integrity-an-essential-area-for-better-peer-review, 12 September 2019

Graf, C. (2019, March 26). Usefulness: An Essential Area for Better Peer Review. Retrieved from https://www.wiley.com/network/researchers/being-a-peer-reviewer/usefulness-an-essential-area-for-better-peer-review, 12 September 2019

Graf, C. (2019, April 2). Timeliness: An Essential Area for Better Peer Review. Retrieved from https://www.wiley.com/network/researchers/latest-content/timelines-an-essential-area-for-better-peer-review, 12 September 2019

Moher D. (2018). Reporting guidelines: doing better for readers.*BMC Med* . 16(1):233. http://dx.doi.org/10.1186/s12916-018-1226-0

Ortuzar N. (2019, June 5) Practicing Diversity and Inclusivity: An Editor's Role. Retrieved from https://www.wiley.com/network/journaleditors/editors/practicing-diversity-and-inclusivity-an-editor-s-role, 13 September 2019

# SUPPLEMENTARY INFORMATION

| Journal Code | Subject Category | Ownership |
|---|---|---|
| J1 | Health Sciences | Society |
| J2 | Physical Sciences & Engineering | Proprietary |
| J3 | Life Sciences | Proprietary |
| J4 | Health Sciences | Proprietary |
| J5 | Health Sciences | Proprietary |
| J6 | Health Sciences | Proprietary |
| J7 | Life Sciences | Proprietary |
| J8 | Life Sciences | Proprietary |
| J9 | Life Sciences | Proprietary |
| J10 | Social Sciences & Humanities | Proprietary |
| J11 | Health Sciences | Society |
| J12 | Social Sciences & Humanities | Proprietary |
| J13 | Social Sciences & Humanities | Society |
| J14 | Social Sciences & Humanities | Society |
| J15 | Social Sciences & Humanities | Joint |
| J16 | Social Sciences & Humanities | Proprietary |
| J17 | Health Sciences | Society |
| J18 | Social Sciences & Humanities | Society |

| Journal Code | Subject Category | Ownership |
|---|---|---|
| J19 | Social Sciences & Humanities | Society |
| J20 | Life Sciences | Society |
| J21 | Social Sciences & Humanities | Society |
| J22 | Health Sciences | Society |
| J23 | Health Sciences | Society |
| J24 | Health Sciences | Society |
| J25 | Health Sciences | Society |
| J26 | Social Sciences & Humanities | Society |
| J27 | Health Sciences | Society |
| J28 | Life Sciences | Society |
| J29 | Life Sciences | Society |
| J30 | Health Sciences | Society |
| J31 | Health Sciences | Proprietary |
| J32 | Health Sciences | Proprietary |
| J33 | Life Sciences | Proprietary |
| J34 | Health Sciences | Society |
| J35 | Life Sciences | Joint |
| J36 | Life Sciences | Society |
| J37 | Health Sciences | Proprietary |
| J38 | Life Sciences | Proprietary |
| J39 | Life Sciences | Proprietary |
| J40 | Life Sciences | Society |
| J41 | Health Sciences | Proprietary |
| J42 | Health Sciences | Society |
| J43 | Health Sciences | Proprietary |
| J44 | Life Sciences | Proprietary |
| J45 | Social Sciences & Humanities | Society |
| J46 | Life Sciences | Proprietary |
| J47 | Physical Sciences & Engineering | Society |
| J48 | Health Sciences | Proprietary |
| J49 | Life Sciences | Society |
| J50 | Health Sciences | Proprietary |
| J51 | Social Sciences & Humanities | Proprietary |
| J52 | Social Sciences & Humanities | Proprietary |
| J53 | Life Sciences | Proprietary |
| J54 | Life Sciences | Proprietary |
| J55 | Health Sciences | Proprietary |
| J56 | Health Sciences | Joint |
| J57 | Social Sciences & Humanities | Proprietary |
| J58 | Health Sciences | Proprietary |
| J59 | Physical Sciences & Engineering | Proprietary |
| J60 | Physical Sciences & Engineering | Proprietary |
| J61 | Life Sciences | Society |
| J62 | Health Sciences | Society |
| J63 | Health Sciences | Proprietary |
| J64 | Health Sciences | Society |
| J65 | Life Sciences | Society |
| J66 | Physical Sciences & Engineering | Society |
| J67 | Social Sciences & Humanities | Proprietary |
| J68 | Social Sciences & Humanities | Proprietary |

| Journal Code | Subject Category | Ownership |
|---|---|---|
| J69 | Life Sciences | Society |
| J70 | Health Sciences | Proprietary |
| J71 | Physical Sciences & Engineering | Society |
| J72 | Social Sciences & Humanities | Proprietary |
| J73 | Life Sciences | Proprietary |
| J74 | Life Sciences | Proprietary |
| J75 | Social Sciences & Humanities | Proprietary |
| J76 | Health Sciences | Society |
| J77 | Health Sciences | Proprietary |
| J78 | Health Sciences | Proprietary |
| J79 | Social Sciences & Humanities | Proprietary |
| J80 | Physical Sciences & Engineering | Society |
| J81 | Physical Sciences & Engineering | Society |
| J82 | Social Sciences & Humanities | Proprietary |
| J83 | Physical Sciences & Engineering | Society |
| J84 | Life Sciences | Proprietary |
| J85 | Social Sciences & Humanities | Proprietary |
| J86 | Social Sciences & Humanities | Society |
| J87 | Social Sciences & Humanities | Society |
| J88 | Life Sciences | Proprietary |
| J89 | Health Sciences | Joint |
| J90 | Health Sciences | Society |
| J91 | Social Sciences & Humanities | Proprietary |
| J92 | Health Sciences | Proprietary |
| J93 | Life Sciences | Society |
| J94 | Life Sciences | Society |
| J95 | Health Sciences | Society |
| J96 | Life Sciences | Proprietary |
| J97 | Life Sciences | Proprietary |
| J98 | Physical Sciences & Engineering | Proprietary |
| J99 | Physical Sciences & Engineering | Proprietary |
| J100 | Health Sciences | Society |
| J101 | Life Sciences | Joint |
| J102 | Social Sciences & Humanities | Joint |
| J103 | Social Sciences & Humanities | Society |
| J104 | Social Sciences & Humanities | Proprietary |
| J105 | Social Sciences & Humanities | Proprietary |
| J106 | Health Sciences | Proprietary |
| J107 | Health Sciences | Proprietary |
| J108 | Health Sciences | Society |
| J109 | Health Sciences | Proprietary |
| J110 | Health Sciences | Proprietary |
| J111 | Life Sciences | Society |
| J112 | Social Sciences & Humanities | Society |
| J113 | Health Sciences | Proprietary |
| J114 | Health Sciences | Society |
| J115 | Health Sciences | Joint |
| J116 | Social Sciences & Humanities | Society |
| J117 | Life Sciences | Society |
| J118 | Physical Sciences & Engineering | Society |

| Journal Code | Subject Category | Ownership |
|---|---|---|
| J119 | Health Sciences | Proprietary |
| J120 | Health Sciences | Society |
| J121 | Life Sciences | Society |
| J122 | Physical Sciences & Engineering | Society |
| J123 | Health Sciences | Society |
| J124 | Health Sciences | Proprietary |
| J125 | Health Sciences | Society |
| J126 | Life Sciences | Joint |
| J127 | Physical Sciences & Engineering | Joint |
| J128 | Social Sciences & Humanities | Proprietary |
| J129 | Health Sciences | Proprietary |
| J130 | Social Sciences & Humanities | Proprietary |
| J131 | Health Sciences | Proprietary |
| J132 | Social Sciences & Humanities | Proprietary |

**Supplementary Table 1:** Journals by subject area and ownership.

| Difference | Number |
|---|---|
| 2 | 58 |
| 1 | 133 |
| 0 | 3796 |
| -1 | 1735 |
| -2 | 89 |
| 'N/A' or 'U' | 525 |

**Supplementary Table 2:** Numbers of overall differences between SA-scores and R-scores (e.g. an SA-score of 2 correlated with an R-score of 1, meaning a difference of -1).

| Subject category | Mean SA-Score | Mean R-Score | difference |
|---|---|---|---|
| Health Sciences | 2.22 | 1.92 | 0.29 |
| Life Sciences | 2.09 | 1.78 | 0.31 |
| Physical Sciences & Engineering | 2.32 | 1.83 | 0.49 |
| Social Sciences & Humanities | 2.14 | 1.90 | 0.24 |
| **Overall** | **2.17** | **1.87** | **0.30** |
| *Range (lowest to highest)* | *0.23* | *0.14* | |

**Supplementary Table 3:** Mean SA-scores and R-scores by subject category, ranked by difference.

| Subject category | Number | % |
|---|---|---|
| Social Sciences & Humanities | 59 | 3.72% |
| Life Sciences | 37 | 2.27% |
| Health Sciences | 32 | 1.31% |
| Physical Sciences & Engineering | 12 | 1.79% |

**Supplementary Table 4:** Number of 'N/A' responses by subject category.