10

Is Peer Review Fit for Purpose?

Malcolm Tight

Introduction

Peer review is endemic to judgement in higher education, as well as throughout the social world. Indeed, it should not be surprising to learn that its origins—like those of higher education in general—are religious, involving the pre-publication judgement on whether a book should be permitted to be put on sale or burnt, perhaps along with its author, as heretical (Lipscombe, 2016).

In contemporary higher education, it is widely assumed that when we need to make a judgement on the quality of something—for example, a student's performance, the employment or promotion of an academic, the importance of an academic publication, whether someone should get a research grant—then we may rely on the assessment of one, two or multiple peers, typically but not always more senior academics. While available quantitative, and thus seemingly more 'objective', data also has

M. Tight (✉)

Lancaster University, Lancaster, UK

e-mail: m.tight@lancaster.ac.uk

an increasing role to play in these assessments—for example, in the context of the four examples given, course grades, student evaluations of teaching, numbers of citations, and previous grants held, respectively—the final judgement will typically be taken by a small number of academics (and increasingly perhaps professional administrators or managers).

This chapter will illustrate and challenge the assumptions underlying peer review, and assess how ‘fit for purpose’ it is in twenty-first century mass higher education. The chapter will focus on different practices of peer review in the contemporary higher education system as practiced by academics (i.e. it will not consider peer review between students, an increasingly popular means of both developing student skills and reducing academics’ workloads). It will question as to how well they work, how they might be improved and what the alternatives are. While the presentation will be grounded in the UK experience, and in English language publications, the discussion will spread out internationally and comparatively.

Three main examples of academic peer review will be introduced and discussed: the refereeing of academic journal articles, the assessment of doctoral degrees and the UK Research Excellence Framework (REF). These have been chosen not simply for their broad and international significance—this is self-evident in the case of journal articles and doctoral degrees, while the UK REF and its predecessor have served as models elsewhere—but because the author has a lot of experience in each of them. Thus, in the first example I have been a widely published author and, across my career, an editor of multiple journals. In the second example, I have been involved, as research supervisor or examiner, in well over 100 doctoral degree examinations; chiefly in the UK, but (mainly at a distance) in a number of other countries as well. In the last example, I served as a member of the Education Sub-Panel in the 2014 REF exercise.

Following a brief discussion of the methodology adopted in this chapter, I will proceed by considering each of my three examples in turn, before drawing some more general conclusions.

Methodology

The methodology adopted in this chapter may be characterized as being informed by systematic review and personal experience. While this may seem a somewhat odd combination, it is an approach that I have refined over the years: you focus on something of interest and explore all of the existing research that you can access and analyse.

Systematic review (Torgerson, 2003) principles have been used to identify relevant published articles on the topics discussed, using keywords and relevant online databases (chiefly Scopus and Google Scholar). However, rather than conduct a full systematic review—which a chapter of this length doesn't really allow scope for—an informed selection has then been made from among the thousands of articles identified for discussion in this chapter.

Personal experience has been used, as already indicated, both in the selection of the topics to be discussed, and in knowing where to look for useful evidence. It has also then, naturally enough, underlain the critique presented. Given the mixture of systematic review and personal experience, the discussion necessarily mixes the third and first persons.

No new empirical data is, therefore, presented or used in this chapter. Rather it rests on the accumulation and interrogation of evidence from past research into the topics of interest. It is, thus, also an example of documentary research (Tight, 2019a).

Refereed Journal Articles

Nearly 20 years ago, with many years' experience of the journal publication process—both as an author and an editor—already under my belt, I decided to carry out a personal test of the veracity of the article reviewing process (Tight, 2003). I had kept copies of all of the reviews of the articles I had submitted to journals over the previous ten years, together with the decisions taken on them by the editors concerned and the comments made by referees.

I went through each article, assessing whether the reviews were positive or negative in tone, or a mixture of the two. While this was, admittedly, a subjective assessment, it was surprisingly easy to do. I then cross-tabulated the results against the editors' verdicts, which (analogous to the PhD examination process discussed in the next section) were typically one of four decisions: accept, minor revisions, major revisions or reject.

The pattern this exercise revealed was quite striking: the relationship between referees' ratings and editorial decisions was far from clear. Highly criticized articles had sometimes been accepted with little or no amendment required, while positively reviewed articles were sometimes rejected. Where one or more referees were positive, and one or more negative, the editorial decision might, of course, go either way.

Clearly, then, the opinion of one's peers was not the only factor that mattered—other considerations were also in play. From personal editorial experience, I would say that two of these additional factors are the editor's own opinions (also a form of peer review of course) and the limitations imposed by the amount of publication space available in the journal (i.e. some editors are looking for reasons to reject articles, while other editors are looking for reasons to accept them), but there are doubtless other factors as well.

Perhaps unsurprisingly, this topic has also been the subject of more extensive, and less personal, research. It would be strange if such a central aspect of academic life had not attracted such attention:

Authors, manuscripts, reviewers, journals and readers have been [sic] scrupulously examined for their qualities and competencies, as well as for their "biases", faults or even unacceptable behavior. This trend has risen with the pioneering work of Peters and Ceci (1982) who resubmitted to journals articles that they had already published, simply replacing the names of the authors and their institutions with fictitious names and making minor changes to the texts. Much to their surprise, almost all of the manuscripts were rejected, and, three exceptions aside, without any accusation of plagiarism. Thirty years later, hundreds of studies on manuscript evaluation are now available. The diverse arrangements of manuscript evaluation are thus themselves systematically subjected to evaluation procedures. For example, in order to comparatively value single blind and double blind, studies have increasingly used randomized controlled trials, leading to opposite results and recommendations for journal editors. (Pontille & Tornø, 2015, p. 75)

Many contemporary studies have focused on the experience of a specific journal or nation. Thus, Hewings (2004) analysed 228 reviews submitted to the journal *English for Specific Purposes*, finding that ‘reviewers take on multiple roles, at the same time discouraging the publication of work that fails to meet the required standards and offering encouragement to authors and guiding them towards publication’ (p. 247). Atjonen (2018) surveyed the opinions of 121 Finnish researchers on the ethics of peer review, concluding that:

Out of nine ethical principles honesty, constructiveness, and impartiality were appreciated but promptness, balance, and diplomacy were criticized. According to two open questions, a third of authors praised and blamed reviewers as experts and non-experts. The accuracy of feedback was more often present in the best rather than in the worst experienced review processes. Journals’ editors and their decision-making called forth more negative than positive accounts. (p. 359)

In a third example, Falkenberg and Soranno (2018) analysed 49 reviews of 26 submissions to *Limnology and Oceanography: Letters*. They found that ‘editor perception of review quality was based on review content rather than if there was agreement on the manuscript decision’ (p. 1), which is somewhat reassuring.

Journal article peer review has also been the subject of large-scale research synthesis. Bornmann et al. (2010) undertook a meta-analysis of studies of the journal peer review process. This involved identifying all of the previous quantitative studies of the topic which they could, and combining their data, focusing on the inter-rater reliability of reviewers (i.e. the extent to which article reviewers make the same recommendations). They identified 70 reliability coefficients from 48 studies, which together had examined the assessment of 19,443 manuscripts. They found that the inter-rater reliability was low; that is, journal reviewers seldom agreed with each other. Meta-regression analyses found that neither discipline nor the method of blinding (i.e. anonymizing) manuscripts impacted on this result.

This might, of course, be at least partly explained by the relatively simple rating scale used by many journals; the accept/major/minor/reject scale already referred to. What constitutes major revisions to one reviewer

might, for example, easily be called minor revisions by another. There are also, however, the occasional cases where one reviewer recommends that an article should be accepted without any further work, and another reviewer recommends rejection. While the obvious response of seeking a third opinion (perhaps that of the editors themselves) is pragmatic, it does ignore the underlying disparity of judgement.

These analyses suggest that both the practices of peer review of academic journal articles and the accuracy of its results may be challenged. Editors, of course, would probably not want to encourage too much of this: the editorial role is demanding enough as it is.

Another response, however, is to question just how much this matters. After all, authors receiving reviews of their work—even when it is rejected by the journal in question—are hopefully receiving at least some useful information, which they may use in revising their articles for possible publication elsewhere. There are usually many alternative journals available, with higher or lower acceptance thresholds, in which publication may be sought. Academic authors simply have to get used to the rough and tumble of the article publication process (to which they themselves contribute as reviewers) if they are to succeed.

It is also possible, in certain circumstances—including that the authors concerned have a strong sense of the worthwhileness of their work, which should come with experience, and are able to defend or respond to criticisms of it—for the authors to negotiate with journal editors, and even through the editor with their reviewers, over the treatment of their submission after a decision has been taken on it (Kumar et al., 2011). This can work to mutual benefit. Thus, in their study of selected science and engineering articles, Kumar et al report that:

Most types of negotiations helped authors to improve presentation of their underlying concepts, quality, clarity, readability, grammar and technical contents of the article, besides offering an opportunity to rethink about several other aspects of the article that they overlooked during the preparation of manuscript. (p. 331)

It is, of course, unrealistic in any case to expect unanimity of judgement (the closest we might get to ‘objectivity’) amongst academics. Some

may warm to a particular line of argument, theoretical framework and/or methodology, while others will be put off by it. The academic world, at least in research terms, is built to a large extent on competition and disagreement (a brief visit to any academic conference should confirm this). To some extent, reviewers might also be said to be acting in a 'zero-sum' game; that is, if they recommend the rejection of an article they are reviewing, there is potentially that much more space available for their own publications.

It would be hard, however, to argue that the academic journal article review process works well, for, in addition to taking up an inordinate amount of (typically unpaid) time and effort, it causes a great deal of emotional upset among those whose efforts are being judged. It may be, of course, that the growing moves towards online and freely available publication, and towards researchers self-publishing their articles on their own websites, will go some way towards resolving these issues.

The Assessment of Doctoral Degrees

Over my career I have been involved in well over 100 doctoral degree examinations: initially my own (which was a disaster), and then as internal examiner, external examiner, supervisor and independent chair, and even as a third examiner brought in to adjudicate between the first two. While the vast majority of these examinations or vivas have been in the UK, I have also participated at a distance in several doctoral examinations in Australia and Hong Kong, and witnessed them (as a member of the public) in Finland and Sweden.

My own doctoral examination was an interesting induction into this experience: my viva lasted for 20 months! I had two external examiners, rather than the more typical UK pattern of an internal examiner (from the candidate's department) and an external examiner, because my department was then trialling measures for making doctoral examinations more robust. Unfortunately, however, these two external examiners could not agree on their recommendations, and it took the department 20 months to bring them to an agreement. As the university I was studying with then had no regulations for dealing with these circumstances, as

universities typically did not then (when the idea of the student as customer had yet to take hold), my viva technically just went on and on.

Since that induction, I have experienced a wide range of viva experiences: ones where the candidate cried, one where the candidate tried to hit one of the examiners, one where the candidate was invited to contribute to a forthcoming book being edited by one of the examiners, one where the candidate was failed for plagiarizing the entire thesis, one where the candidate failed to give a straight answer to any of the examiners' questions, even one where I fell asleep (I was the non-participating attendant supervisor for that one)! However, the vast majority, if sometimes somewhat underwhelming, have led—usually after major or minor revisions, but occasionally awarded straight away—to a doctoral degree.

While it seems clear that the doctoral examination process in the UK works—if it did not, there would have been increasingly strident calls for change before now—this is not to say that it works well and consistently. Even within a single country, there is considerable variation in institutional and disciplinary practice (Tinkler & Jackson, 2000), and a variety of doctoral models. In the last few decades, professional or taught doctorates have become popular, alongside the traditional format of a lengthy thesis produced by an individual after a few years of supervised research, changing the dynamic and expectations.

Examiners can vary a great deal in the attention that they give to a thesis. One examiner may produce a report of 20, closely typed, pages, while another may turn in a single paragraph. And, of course, how one examiner interprets major or minor revisions may be very different to another. A lot may actually come down to whether the examiner in question 'likes' the candidate and topic.

The scope for variation was demonstrated to me clearly when a candidate at another university appealed against the examiners' decision (which was 'major revisions'), winning their case on what might be termed a technicality. Another viva was held with two new examiners, and, while they also returned a verdict of 'major revisions', the viva was 'friendlier' and the revisions were much more achievable (as well as being different from the first set). While the notion that the 'academic judgement' of the examiners chosen, as opposed to the examination process, cannot be challenged is still upheld, it is clearly coming under some threat.

Whether the verdict of two academics on a doctoral thesis—widely seen nowadays as the entry point to an academic career—can be relied upon, therefore, is highly debatable, even if the doctorate is only seen as setting a certain minimum standard. Much clearly depends critically on which two—or which panel of—academics are picked as examiners. I now warn all of my research students that they might be lucky and get the only two academics in the world who would be prepared to pass their thesis as their examiners, or they might be unlucky and get the only two who would fail it. This may be an exaggeration, but only slightly, and it certainly does not make the task—typically born by the candidate's supervisor—of choosing examiners any easier.

Some authors have argued that it is time, at least in the UK:

for a radical review of doctoral education assessment across disciplinary boundaries to consider systematic and universally agreed criteria and scrutiny procedures to quality assure the award. For example, measures such as the convening of a public panel for the viva on the continental model are worthy of consideration, with this open forum removing the secrecy element from the process. With an increase in work-based, professional and 'taught' doctorates, a further aim of such a review would be to develop standardised procedures across disciplines and institutions to 'benchmark' standards in both the written and oral assessment components. Also, the establishment of codes of practice concerning examiner selection that, for example, might move towards the appointment of anonymous reviewers and the mandatory nomination of an independent chair for the viva, might increase confidence in the integrity of the doctoral assessment process. Whilst supervisors may perceive such policies as a threat to their academic autonomy and thus might resist their implementation, they may help to positively transform the current disparities through which inequalities and inconsistencies are maintained. (Watts, 2012, pp. 379–380)

One does, however, have to question just how realistic some of these recommendations are, and also whether Watts' understanding of 'the continental model' is as complete as it might be. To expect 'universal agreement' 'across disciplinary boundaries' is, after all, asking rather a lot of a disparate professional group for whom the favourite managerial metaphor is 'herding cats'! On the other hand, 'standardised procedures' and

‘codes of practice’ do already exist, at least at a disciplinary level, but the point is that they remain subject to individual interpretation.

Of course, as Watts recognizes, practices regarding the assessment of doctoral degrees also vary significantly from country to country. In some, including many European countries, it is a public event, but the result is pre-determined in private beforehand (van der Heide et al., 2016). In other countries, such as in North America, it is a committee decision and a viva may not be held. In some countries, the doctorate is even graded, such that only those who pass with a certain grade are eligible for academic appointments.

The extent to which the more labour intensive of these practices, however ‘good’ they might be, can survive as doctoral education—like the rest of higher education—increasingly becomes a mass market, is, though, debatable. In addition to consistency and transparency, the assessment of doctoral degrees needs to be time efficient.

Again, though, as for the refereeing of journal articles, it might be questioned how much the deficiencies identified in the doctoral examination process really matter. The vast majority of students who work hard on their research and thesis over a period of years, with their supervisor’s support, are awarded a doctoral degree, usually after some more work following their viva (i.e. they are recommended to undertake minor or major revisions). The doctorate is ‘a PhD, not a Nobel Prize’ (Mullins & Kiley, 2002), an indication that the candidate is judged fit to undertake independent research in their own field, and to supervise others. It is merely an early staging post in the academic ‘journey’, not its end.

The UK Research Excellence Framework (REF)

The UK REF—and its predecessor the Research Assessment Exercise (RAE)—is a particularly high-stakes exercise, determining a large part of the research funding received by universities and their component departments. Despite calls to make use of available metrics (such as citation rates), its judgements remain based on peer review by ‘expert’ panellists (Koya & Chowdury, 2017; Marques et al., 2017; Mryglod et al., 2013).

From personal experience on the Education Sub-Panel in 2014, I can confirm that this involved a great deal of reading, discussion, benchmarking, cross-checking and debate. But it was not as onerous a task as some would have you believe. After all, as a supposed expert on a particular field, you were already familiar with the work of many of its researchers, and had already read—but in a different, less judgmental, context—many of the outputs that you were now required to rate.

Curiously, the details of this expensive and time-consuming exercise are not made available. All that the submitting institutions receive is a brief summary for each unit of assessment. This means that, at best, they can only ‘second guess’ what individual ratings led to the overall ratings given, and that they may then base their future planning on erroneous interpretations.

The relevance of the REF and the RAE to UK higher education is obvious, but it has also been influential in many other countries which have adopted their own versions. These countries include, for example, Australia, China, Estonia, Hong Kong, Ireland, Italy and Japan.

Not surprisingly, since they were designed to assess and reward the research prowess of all UK academics and their employing institutions—or, at least, those research-active academics that their universities chose to submit—the RAE and REF have been the subject of a great deal of speculative, critical and evaluative research by UK (and other) academics. Some of these studies will be quoted here to illustrate the range and depth of the critique.

Interestingly, for the purposes of the present chapter, Bence and Oppenheim (2004) drew the analogy between peer review, as practiced for journal articles, and the peer review of submitted outputs (typically published journal articles) undertaken by the subject panels of experts set up for the RAE. They argued that the secondary assessment of articles that had already passed peer review was poor practice:

The links between the RAE, journal peer review and quality are complex. The use of peer review for refereeing papers submitted for publication has evolved to become a self-policing mechanism for the community, by the community, which attempts to maintain quality standards and to an extent guard the reputation of journals... the academics doing the judging [in the RAE] are from other

institutions in the same sector, essentially competing for the same resources, and yet are relying on secondary subjective judgements of earlier peer-review decisions. This would be fine if everyone trusted the outcomes of peer review; but they do not. We conclude that because of the many criticisms of peer review, it may be unwise to base funding decisions on second level peer review of articles that have already undergone initial peer review. (pp. 363–364)

Others addressed the issue of the apparent improvement in research ratings revealed by the RAE over time and attempted to explain this. For example, Sharp (2004) examined the RAE results for the three successive exercises of 1992, 1996 and 2001, focusing on differences between years and between units of assessment (i.e. subjects or disciplines):

The results show that mean ratings have improved markedly over time, particularly between 1996 and 2001, but that this upward shift is unevenly spread across Units of Assessment. In both 1996 and 2001, mean ratings varied significantly across Units of Assessment, with higher means being associated with Units in which there were fewer submissions. (p. 201)

While Sharp was very careful and measured in his comments and conclusions, it seems clear that—given that the assessment panels were recruited from the departments being assessed, and particularly when the panels were relatively small—the possibilities for some, perhaps unconscious, inflation in ratings given, so as to protect or enhance the relative standing of one's own discipline, were there.

Others could afford to be rather more overt in their critique of what became widely derided as 'game playing', but might simply represent pragmatic institutional decision-making designed to maximize their RAE results and the financial benefits that followed them. Thus, Moed (2008), based in the Netherlands, was able to persuasively chart how institutional strategies changed over time to respond to the continual changes made to the RAE's methodology:

A longitudinal analysis of UK science covering almost 20 years revealed in the years prior to a Research Assessment Exercise (RAE 1992, 1996 and 2001) three distinct bibliometric patterns, that can be interpreted in terms of scientists' responses to the principal evaluation criteria applied in a RAE. When in the

RAE 1992 total publications counts were requested, UK scientists substantially increased their article production. When a shift in evaluation criteria in the RAE 1996 was announced from 'quantity' to 'quality', UK authors gradually increased their number of papers in journals with a relatively high citation impact. And during 1997–2000, institutions raised their number of active research staff by stimulating their staff members to collaborate more intensively, or at least to co-author more intensively, although their joint paper productivity did not. This finding suggests that, along the way towards the RAE 2001, evaluated units in a sense shifted back from 'quality' to 'quantity'. (p. 153)

Alongside the pervading critique of the RAE and REF as a grotesque game, the purpose of which was to deliver the lion's share of the available research funding to the oldest and best-established universities and departments, the most prevalent critique, however, has probably been about whether peer review was the best way of undertaking the evaluation. This critique has had a number of elements. Thus, in economic terms, it has been argued that the costs of the RAE and REF, principally in terms of the hours of academic and administrative time taken in putting together departmental and institutional submissions, and then in evaluating them, were very hard to justify. After all, would not this time be better spent in actually doing some more research?

There is also a disciplinary element to this critique, however, arguing that bibliometric methods—the obvious alternative to peer review, and used by some RAE and REF panels to inform their decisions—are not appropriate to all disciplines. For example, in the case of social policy and social work, McKay (2011) argued that:

Using quantitative evidence it seems possible to base estimations of research environment on observable data, or at least to regard such data as a valuable check on the assessment. The same may not be said of the evaluation of research outputs, at least in SPA [social policy and administration] and SW [social work], although there are disciplines where journal rankings correlate very strongly with the outcome. (p. 540)

The underlying argument is that in some disciplines, notably the hard sciences—where quantitative methods dominate—the relative standing of particular journals is widely acknowledged and citation rates are

substantial. In other disciplines, however, notably the arts, humanities and social sciences—where qualitative methods are much more popular—there is not such a clear ‘pecking order’ among journals and citation rates are generally low, making bibliometric data a much less useful guide to quality.

This argument is somewhat supported in the Italian context by Abramo et al. (2011), who examined the experience of the first Italian research assessment (the VTR) and its planned replacement (the VQR). They considered the use of bibliometric exercises as a replacement for peer review, arguing that the former, as well as being less time-consuming, would yield a better result:

For the Italian VTR, the objective was to identify and reward excellence: in this work we have attempted to verify the achievement of the objective. To do this we compared the rankings lists from the VTR with those obtained from evaluation simulations conducted with analogous bibliometric indicators... The results justify very strong doubts about the reliability of the VTR rankings in representing the real excellence of Italian universities, and raise a consequent worry about the choice to distribute part of the ordinary funding for university function on the basis of these rankings. One detailed analysis by the authors shows that the VTR rankings cannot even be correlated with the average productivity of the universities. Everything seems to suggest a reexamination of the choices made for the first VTR and the proposals for the new VQR. The time seems ripe for adoption of a different approach than peer review, at least for the hard sciences, areas where publication in international journals represents a robust proxy of the research output, and where bibliometric techniques offer advantages that are difficult to dispute when compared to peer review. (p. 940)

Note the qualifying statement they carefully include—‘at least for the hard sciences’—thus effectively supporting McKay’s position.

In the Irish context, Holland et al. (2016) extend McKay’s argument to the whole of the arts, humanities and social sciences, and lay the blame for what they clearly regard as an unwarranted imposition on the forces of neoliberalism (an oft-chosen target for academics and others at the present time, and one that is sufficiently nebulous not to need to fight back: see Tight, 2019b).

The dynamic of what is being valued within research assessment exercises in higher education in Ireland and elsewhere is changing as a result of the re-

emergence of neoliberalism in the context of the global recessionary economic climate. AHSS [arts, humanities and social sciences] researchers are becoming increasingly concerned at the lack of inclusivity in what is being valued as research outputs, and in what can be counted within research assessment exercises. Evidence is emerging that quantitative metrics are more valued within neoliberal agendas, and that this is changing the behaviour of researchers towards engaging in and disseminating research that can readily contribute to such quantitative metric profiles... More appreciation for the diversity of research, and the appropriate assessment of quality thereof, within AHSS disciplines needs to be fostered within research assessment exercises... Academics and researchers in the Arts, Humanities and Social Sciences need urgently to reach agreement on what should be valued in terms of research activities, outcomes and/or impacts, and at what level (institution, department, unit, or individual). They also need to reach consensus with key policy-makers on how this work can be suitably assessed within the broader context of performance assessment in higher education. (p. 1113)

The obvious weakness of this argument, however, is the evident difficulty which Holland, Lorenzi and Hall have in specifying what might constitute a high quality research output in the arts, humanities and social sciences. Of necessity, therefore, we have to fall back on disciplinary peer review, that is, what do our colleagues and superiors think?

We may, of course—as we did in the two previous sections—again pose the question ‘does it really matter?’ The RAE and REF may not be wholly fair or objective exercises, and they are certainly not transparent. Unlike journal article or doctoral degree peer review, there is also no real scope for negotiation or appeal over the results. But, if we accept that available research funds should be targeted towards those who are making the best, or at least the most, use of them, is there a better system?

Conclusion

In this chapter, we have considered the use of peer review in higher education in the context of the evaluation of journal articles, doctoral degrees and institutional research performance. While the examples have been linked to my own experience and grounded primarily in the UK, their broader relevance and applicability is fairly self-evident.

The underlying question driving the discussion in the chapter has been ‘is peer review fit for purpose?’, and to this have been added the related questions of ‘what alternatives are there?’ and ‘does it really matter?’

I think we have to conclude that peer review, being of long standing and fundamental to the operation of the academic enterprise, is not going away any time soon. It does have major flaws in that it is subjective and may be manipulated, but that is another way of saying that it is human. We all have preferences and biases, but—all taken together and in the long run—these should more or less even each other out.

Are there better alternatives? Well, it depends upon your perspective, and here we run straight into the qualitative/quantitative debate that has plagued social science research for decades.

In the case of the RAE/REF, it would be perfectly possible to replace peer review with bibliometric analyses—based on journal status and citation counts—which could be completed much quicker and much more cheaply. This would probably produce not too dissimilar results to peer review for the hard sciences and related fields (i.e. the disciplines that absorb the great majority of research funds). It is doubtful that this would work, or work so well, in the arts, humanities and social sciences, however, and even in the hard sciences something would be lost in terms of the appreciation of the overall field of research.

In the case of both journal article evaluation and doctoral degree assessment, a better alternative is not so clear. It would, of course, be possible to try and improve current practices: through, for example, more careful selection of article reviewers and doctoral examiners, more extensive training and the provision of more written guidance. But this would be to add significantly to the workload of those involved, who are undertaking tasks which, whether we like it or not, are really marginal to their employment, and which are either unpaid or poorly paid.

So, does it really matter? Well, obviously, yes, or I wouldn’t have written this chapter. Peer review is of critical importance to the practice of higher education. It is vital that we perform it as well as we can, bearing in mind its actual and potential deficiencies. We cannot assume that all academics are already competent at it, in its various forms, but need to provide appropriate training, guidance and support. We also need to

keep a watching eye on the results of peer review, allow them to be challenged and be prepared to challenge them ourselves where we believe this to be necessary.

References

- Abramo, G., D'Angelo, C., & Di Costa, F. (2011). National Research Assessment Exercises: A comparison of peer review and bibliometrics rankings. *Scientometrics*, 89, 929–941.
- Atjonen, P. (2018). Ethics in peer review of academic journal articles as perceived by authors in the educational sciences. *Journal of Academic Ethics*, 16(4), 359–376.
- Bence, V., & Oppenheim, C. (2004). The influence of peer review on the research assessment exercise. *Journal of Information Science*, 30(4), 347–368.
- Bornmann, L., Mutz, R., & Daniel, H.-D. (2010). A reliability-generalization study of journal peer reviews: A multilevel meta-analysis of inter-rater reliability and its determinants. *PLoS One*, 5(12), e14331.
- Falkenberg, L., & Soranno, P. (2018). Reviewing reviews: An evaluation of peer reviews of journal article submissions. *Limnology and Oceanography Bulletin*, 27(1), 1–5.
- Hewings, M. (2004). An 'Important Contribution' or 'Tiresome Reading'? A study of evaluation in peer reviews of journal article submissions. *Journal of Applied Linguistics*, 1(3), 247–274.
- Holland, C., Lorenzi, F., & Hall, T. (2016). Performance anxiety in academia: Tensions within research assessment exercises in an age of austerity. *Policy Futures in Education*, 14(8), 1101–1116.
- Koya, K., & Chowdury, G. (2017). Metric-based versus peer-reviewed evaluation of a research output: Lessons learnt from UK's national research assessment exercise. *PloS One*, 12(7), e0179722.
- Kumar, P., Rafiq, I., & Imam, B. (2011). Negotiation on the assessment of research articles with academic reviewers: Application of peer-review approach of teaching. *Higher Education*, 62, 315–332.
- Lipscombe, T. (2016). Burn this article: An inflammatory view of peer review. *Journal of Scholarly Publishing*, 47(3), 284–298.
- Marques, M., Powell, J., Zapp, M., & Biesta, G. (2017). How does research evaluation impact educational research? Exploring intended and unintended

- consequences of research assessment in the United Kingdom, 1986–2014. *European Educational Research Journal*, 16(6), 820–842.
- McKay, S. (2011). Social policy excellence: Peer review or metrics? Analysing the 2008 research assessment exercise in social work and social policy and administration. *Social Policy and Administration*, 46(5), 526–543.
- Moed, H. (2008). UK Research Assessment Exercises: Informed judgments on research quality or quantity? *Scientometrics*, 74(1), 153–161.
- Mryglod, A., Kenna, R., Holovatch, Y., & Berche, B. (2013). Comparison of a citation-based indicator and peer review for absolute and specific measures of Research Group Excellence. *Scientometrics*, 97(3), 767–777.
- Mullins, G., & Kiley, M. (2002). 'It's a PhD, not a Nobel Prize': How experienced examiners assess research theses. *Studies in Higher Education*, 27(4), 369–386.
- Peters, D., & Ceci, S. (1982). Peer-review practices of psychological journals: The fate of published articles, submitted again. *Behavioral and Brain Sciences*, 5(2), 187–195.
- Pontille, D., & Torny, D. (2015). From manuscript evaluation to article valuation: The changing technologies of journal peer review. *Human Studies*, 38(1), 57–79.
- Sharp, S. (2004). The research assessment exercises 1992–2001: Patterns across time and subjects. *Studies in Higher Education*, 29(2), 201–218.
- Tight, M. (2003). Reviewing the reviewers. *Quality in Higher Education*, 9(3), 295–303.
- Tight, M. (2019a). *Documentary research in the social sciences*. Sage.
- Tight, M. (2019b). The neoliberal turn in higher education. *Higher Education Quarterly*, 73(3), 273–284.
- Tinkler, P., & Jackson, C. (2000). Examining the doctorate: Institutional policy and the PhD examination process in Britain. *Studies in Higher Education*, 25(2), 166–180.
- Torgerson, C. (2003). *Systematic reviews*. Continuum.
- van der Heide, A., Rufas, A., & Supper, A. (2016). Doctoral dissertation defenses: Performing ambiguity between ceremony and assessment. *Science as Culture*, 25(4), 473–495.
- Watts, J. (2012). Preparing doctoral candidates for the viva: Issues for students and supervisors. *Journal of Further and Higher Education*, 36(3), 371–381.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.