THE INFLUENCE OF PEER REVIEW ON THE RESEARCH ASSESSMENT EXERCISE

Valerie Bence

and

Charles Oppenheim*
Department of Information Science,
Loughborough University
Loughborough,
Leics LE11 3TU

***To whom correspondence should be addressed**

Abstract.

The use of peer review within both the scholarly communication system and the UK's Research Assessment Exercise is reviewed. The common denominator is that of peer-reviewed academic journals, since peer review is used by referees to aid publication decisions and by RAE panel members to evaluate a department's research performance. We propose that since academic research is now subject to peer review at all stages of evaluation, it is becoming an accepted method of rewarding (by funding) research. The growth of electronic publications (both toll-access and open access) provides possibilities for changes to some of the process of peer review and RAE, but the fundamental model of peer review to reduce the number of poor quality publications will remain. The paper concludes that because of the many criticism of peer review, it is unwise to base funding decisions on second level peer review of articles that have already undergone peer review.

1. Introduction

Peer review of research is difficult to define precisely and encompasses a variety of activities. Gillette [1] suggests that "peer review" is a generic term, which includes impressionistic peer review (which provides an overall judgement), professional-performance peer-review, grant giver peer review and journal article peer review (the main focus of this work). It is implicit in many areas of scholarly activity, e.g., the submission of proposed papers to conferences, the publishing of research monographs, the awarding of research grants/contracts, and the publication of journal articles. The Research Assessment Exercise (hereafter RAE), being a type of peer reviewed assessment exercise, mirrors in some ways the peer review process immured within scholarly publication. By 2002, this type of review was stated as being "the process used to determine how science funding is allocated, which research is published and where it is published" [2]. This article considers both journal publishing and the RAE; the term "peer review" is used here as meaning a quality control method using independent experts applied to both processes. However, we recognise that using the same term for both processes could be confusing, and in this paper distinguish the two where necessary.

The first section of this article examines old and new models of scholarly communication, before considering the processes involved in journal peer review. The subjective concept of quality, when applied to journals, leads to decisions based on perceptions of prestige, ranking and hierarchies of them as vehicles for published research. With the RAE driver and the implicit need to publish in the 'best' journals, but with little or no formal guidance as to what these are, the need for indicators of journal 'quality' has resulted in various forms of measurement. The use of citation counts and impact measurement will be outlined, before finally exploring peer review within the context of the RAE itself.

2. Old and new methods of scholarly communication

The four main functions of scholarly communication [3] have long been accepted as being: -

- the dissemination of current knowledge,
- archiving the canonical knowledge base,
- quality control of published information (via peer review),
- assignment of priority and credit for author's work.

All are key to the activities of scholars, academic libraries, publishers and universities. Each stakeholder has a well-defined part in the process, as outlined in Willis' models of scholarly communication [4] shown in Table 1. This model, although not complete [(1)] shows clearly that on the traditional side, universities pay for virtually the whole process, either directly through research funding, or indirectly through journal subscriptions and the hidden costs in library storage and maintenance.

The process is divided into seven steps - from conducting the research to archiving the printed material. This shows that the work is evenly divided – with three functions performed by faculty and three by publishers, with archiving, being the province of the library.

Table 1.
Alternate models of Scholarly Publishing

	TRADITIONAL MODEL			**NEW MODEL incorporating e-publication**		
Research Function	***Done by***	***Paid for by***	***Value added***	***Done by***	***Paid for by***	***Value added***
Conduct research	Faculty	grant/faculty/ university [(2)]	New knowledge	Faculty	grant/faculty/ university	New knowledge
Generate paper	Faculty	Faculty	Knowledge dissemination	Faculty	Faculty	Knowledge dissemination
Gate-keeping	Faculty	Faculty	Quality	Faculty	Faculty	Quality
Publishing	Publisher	subscriber/university/ library	Structure	Web Group	?	Structure
Marketing	Publisher	subscriber/university/ library	Awareness	Web Group	?	Awareness
Distribution	Publisher	subscriber/university/ library	Convenience	Websites	Internet	Convenience
Archiving [(3)] finding	Library	university	Accessibility	Web Gp/ Library	?	Accessibility

Some problems with the traditional system arise because the major stakeholders - faculty (i.e., academic staff) and universities (including the library) on one side and commercial publishers on the other - have different motivations. Research is created by faculty, supported and resourced by universities, who are generally concerned with the free dissemination of

knowledge and would prefer the most open, cost-effective approach possible. Commercial publishers, however, are profit -driven. When authors submit papers to journals, they assign copyright, or at least publication rights, to the publisher, who in turn makes decisions on subscription rates, marketing and distribution. The two "paid for by" columns in Table 1 illustrate this. In the traditional model, faculty and universities directly and indirectly resource the first three steps in the process and then pay <u>again</u> via subscriptions, for access to the information they paid to create in the first place This is true of the system as a whole, but not true at the level of any particular institution, for that particular institution is giving away its own research output, but what it is buying back from publishers is not just its **own** research output, but also the research output of **other** institutions. The upshot is the same: money is being needlessly spent, and access and impact are needlessly being blocked, but it is not quite the simple “buyback” lament that libraries and Universities have been claiming. In the new model, the first three functions remain the same, being completed by faculty just as before, and distribution is accomplished via Internet websites.

It is possible to take issue with the above model in terms of where the responsibility for peer review lies. It is not shown separately as a function (usually the responsibility of publishers to **administer** peer review, as the peers typically review for free in almost all fields) and as the value-added from this process is ‘quality’, it is unclear whether Willis sees this as being part of his so-called gate-keeping function.

One of the distribution methods in the new model – e-journal Websites, highlights two aspects of the process subject to the greatest change - publishing and marketing. Here, Willis shows these functions being performed by "Web Groups" without definition. In reality, these are professional societies, commercial publishers, university presses etc, which provide the editing infrastructure needed to support scholarly publishing as an activity. It also leaves open the question of costs. E-journals available over the Internet may be supported via traditional subscriptions, site licences, pay-per-view or be free, and there is a large body of literature concerned with developing e-publishing models, many of which continue to emphasise the

importance of the peer review role. Changes due to e-publication are discussed later under the review process.

3. Journal peer review

3.1 Previous literature

Despite the growth in importance of the peer review process, it has not been subject to rigorous scientific study until fairly recently. Its earliest use is thought to be in selecting papers for *Philosophical Transactions* in the mid 17th century, by the Royal Society [6]. Publications on various aspects of the process were sporadic until the mid to late 20th century. These studies have been reviewed by Weller [7] and Abel and Newlin [8].

As journals often received more manuscripts than they could publish, the need for selection of the most suitable, and rejection of the rest increased the pressure for seeking the advice of expert reviewers. The evolution into the current procedure used by the academic community has been tacitly accepted for many years. However, some critics argue against it. For example, Nash questions the validity of the process as being "a historically received practice, not a rationally constructed procedure specifically designed to achieve a clearly defined objective. As a practice two interesting features mark it. Firstly, and strangely, it has remained conceptually un-examined. Secondly, any attention paid to it has evidently been from the perspective of professional interests this practice has tended to touch….but there is no evidence that it has ever been conceptualised or subjected to a rigorous and systematic analysis." [9] However, one could argue that it should be the alternatives that need to be tested and benchmarked against the received system, not the received system that needs to prove itself. True peer review is using qualified experts to evaluate the work of fellow experts, with both the author and the referees answerable to a meta-expert, the editor, who is in turn answerable for his or her decisions via the quality and impact of his journal.

In 1986, Rennie, then deputy editor of the *Journal of the American Medical Association* (*JAMA*) wrote: "There seems to be no study too fragmented, no hypothesis too trivial, no

literature citation too biased or too egotistical, no design too warped, no methodology too bungled, no presentation of results too inaccurate, too obscure, and too contradictory, no analysis too self-serving, no argument too circular, no conclusions too trifling or too unjustified, and no grammar and syntax too offensive for a paper to end up in print" [10].Of course, there is a hierarchy of quality presumably correlated with the rigour and selectivity (including the rejection rate) of the peer review; the impact factor too presumably reflects this. The peer-review hierarchy goes all the way down to a virtual vanity press at the bottom (though still nominally "peer-reviewed").

With such criticism, the peer review process itself has become a growth area for research. Lock [11] offered perhaps the first comprehensive critique of the peer review process and raised awareness of the potential biases embedded in the system. In the last few years, several authoritative works [12, 13,14] about the process of scholarly communication have contributed to the growing debates within the academic community surrounding the whole process and how it might evolve.

In 1998, Meadows reviewed the literature to date [15], looking at the efficacy of peer review and its difficulties. More recently, Weller has published her comprehensive examination of almost 1500 published studies on the peer review process between 1945 and 1997 [7]. She concludes that editorial peer review is evolving but has yet to be replaced. She considers alternative models of the editorial process (termed as "emerging", but by now in full use by many journals and publishers), which include fast-track publication, post-publication review, continual updating of content and online posting of non-reviewed content. Weller's survey concentrates mainly on medicine, which is where most research on the refereeing process has been undertaken, perhaps because, as pointed out by Rowland [16], that this is the field where dependable, reliable, quality controlled information can literally be a matter of life and death. As a result, much progress is being made in the study of peer review, especially within the bio-

medical field. There is a regular series of International Congresses on Peer Review in the Biomedical Sciences, which are reported in special themed issues of the *Journal of the American Medical Association (JAMA)* e.g., [17]. However, the practice of peer review covers all disciplines. Peer reviewed publication is about establishing academic status or boosting institutional research ratings. As more papers are generated, competition for inclusion in prestigious or high impact journals becomes ever more intense, resulting in the editorial and peer review systems struggling under the pressure.

It could be suggested that the RAE itself has had an effect on the reviewing system. One publisher recently stated "my impression is that the RAE has increased the amount of material entering the peer review process without increasing the amount of material leaving it, i.e., it has increased the amount of time devoted to refereeing material which will not be accepted for publication thus increasing the 'noise' in the system" [18].

3.2 Advantages and disadvantages

Some of the advantages and disadvantages of the peer review process are shown in Table 2: -

Table 2.

Advantages and disadvantages of peer review

Advantages	Disadvantages
Refereeing allows an author to claim priority and ownership to an idea.	Reviewers can make factually incorrect judgements, and undetected falsification has occurred.
Validation of the author's work - it provides impartial evaluation of manuscripts to weed out flawed and fraudulent research, i.e., acting as gatekeeper to ensure high standards for published research.	Reviewers are not always impartial, and can allow their own opinions to reflect their judgement. Many forms of bias exist, institutional, gender etc. It should be remembered that reviewers are essentially the authors' competitors.
Protection from plagiarism.	It provides opportunities for stealing ideas and plagiarism.
Assurance of authenticity.	No redress for authors if reviewers are truly anonymous.
Gives credence in the community, which may help in obtaining jobs, promotion or funding.	Time delay is the main problem. Double blind reviewing, especially with revisions, can take many months. Thus publications of important research results may be delayed.
Quality assurance for the journal's standards.	Reviewers can disagree on the merits of the same

	paper.
Improves scholarship by ensuring the citing of relevant literature. Even if rejected, feedback to authors is still valuable; constructive criticism and the process of revision improve manuscripts. Editors can direct to appropriate journals if wrongly submitted.	A lack of comments, even if an article is accepted, is not helpful for authors. As this is an unpaid task, it may not be a priority for academics, which in turn compounds delays. An article may be accepted by one journal that has already been rejected by another. This supports either the 'quality' argument or the 'inconsistency' argument or the fact of a quality hierarchy. Perpetuates the feeling of a clique or closed community. New ideas, those outside the mainstream or contradicting established conventional wisdom, may be blocked. Lee and Harley discuss this within the context of Economics [19].

Source: Adapted from: [20] and [21].

Of course, peer review may not be the best selection method for all journals. Whereas many academic journals publish long scholarly papers and maintain an archive, others publish summaries of current research or informal accounts of conferences (e.g., *New Scientist* and *Harvard Business Review*). Here, the time factor is crucial, and the long delays occurring with peer review are a hindrance. Both types of journals are important and fulfil a role in scholarly communication.

Such a labour intensive and lengthy process is also costly. In 1997, Tenopir and King investigated average publication costs for journals produced in the US [22]. They concluded that the average direct cost to provide the first copy of an article is about $2000 (including article reviewing, refereeing, editing, copy-editing, preparation of illustrations and master copies). Indirect costs such as subscription maintenance, marketing, overheads and costs of reproduction such as paper and printing push the amount up to $4000 per article [(4)]. Donovan also studied costs in 1998 [23], concentrating on peer review costs. He took a fairly small sample of journals published by learned societies mainly within the Science, Technology and Medicine (STM) fields. The results are shown in Table 3.

Table 3.

Costs and rejection rates of refereeing papers submitted to eight learned societies

Journal	Submission rate (Papers/year)	Rejection rate (%)	Submitted papers (£ cost per paper)	Accepted papers (£ cost per paper)
A	9000	50	200	400
B	9000	45	33 *	60
C	-	50	60-70	120-140
D	-	50	100	200
E	2100	52	150	288
F	900	65-70	83	237
G	500	50	50	100
H	650	30	146	209

* other editorial costs and overheads excluded

This shows that the overall process is indeed expensive, and that rejection rates remain quite high, which in itself pushes the cost per article up. Donovan also points out that peer review is a critical component in the competition between rival journals, where "good refereeing and editing raises the perceived quality and increases reader appeal, with increased quality comes increased citation of published articles in other works" [23]. The premise is that highly cited journals attract more submissions. It should follow therefore, that quality journals spend more on the refereeing process, with some investment wasted on rejected material (Table 3 shows that the cost per accepted paper increases to allow for rejected papers). Thus, peer review costs form an important part of journal economics. Rowland has also undertaken an in-depth study of associated costs [24]. We therefore do not examine the matter further here.

Despite the disadvantages, peer review retains an essential role in academic publishing, and the development of electronic journals and e-print archives does not affect this. Indeed, the first recommendation approved at the ICSU/UNESCO Conference of Experts on Electronic Publishing in Science (1996) re-iterated that "strict peer review should be applied to all scientific material submitted for publication in e-journals" [23].

In research assessment too, a review published in 2003 (on behalf of the UK higher education funding bodies) [25] confirmed that the RAE is essentially a 'secondary' peer review exercise (secondary, because the articles being assessed have themselves already undergone primary peer review) and the very first recommendation was that "any system of research assessment designed to identify the best research must be based upon the judgement of experts" [26].

Weller states that "the underlying strength of editorial peer review is the concerted effort by large numbers of researchers and scholars who work to assure that valid and valuable works are published, and conversely to assure that the invalid or non-valuable works are not published" [27]. But, as her study (and others) shows, the ability of peer review to "assure" quality is debatable; as a quality control system it does assist, but does not by itself guarantee a valid and reliable outcome.

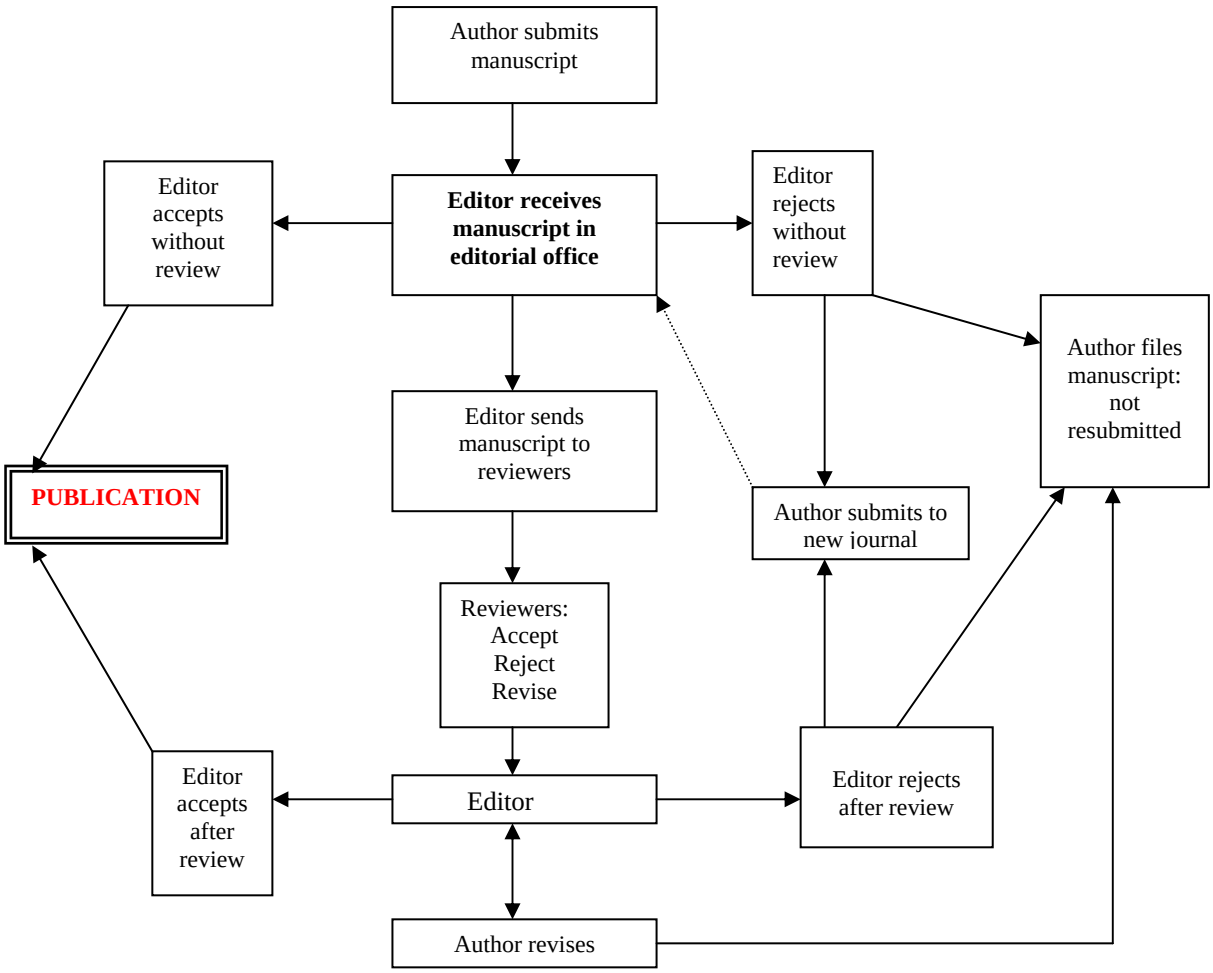
4. The peer review process

4.1 Traditional peer review (for print journals)

There is much variation in the ways in which peer review has become institutionalised at various journals and the process itself is one, which many authors only have a basic understanding of when they submit papers for publication. Chubin describes the process of refereeing submitted articles as "one of negotiation and controlled compromise…an extended trilateral negotiation amongst authors, editors and reviewers". The referee's role in this trilateral negotiation "is to limit the author's claims and shape the editor's judgements"[28]. The editor then uses referees' comments as an aid to deciding whether or not to publish the paper as it stands or at all. Chubin points out a paradox, whereby "the open sharing of knowledge through publication is preceded by secret deliberations among a few handpicked specialists, acting with restricted information, often vague and unenforceable guidelines and with little accountability to authors." [29] However, we would argue that this is mo different to the public verdict of a jury that is preceded by secret deliberations, and closed drug-company testing precedes the open adoption of a drug.

All journals with a preset number of articles per edition are a limited resource. Each article accepted for publication removes a 'publication window' for someone else [30] at the level of that journal. Therefore, competitive market conditions apply and this competition extends to academic researchers looking to place their work in quality journals in their field. This competitive element partly led to the evolution of peer review as a filter for academic publication, and it has become an integral part of scholarly communication. Although it may differ slightly between individual journals, an overview of the traditional process is described below.

Fig. 1. Path of a manuscript through the traditional peer review process



Source: [31].

Figure 1 outlines the stages in the traditional peer review process, from receipt of an article by the Editor, to the end result, publication. Some journals publish articles on editorial acceptance alone, some after review by an Editorial Board, but most that use refereeing use a formal reviewing process, usually open or closed review [(5)].

One argument against closed review is that it can seem unacceptable for judgements to be made on the work of others in secret – but this is what works in other areas of life, e.g., juries, and voting in elections. Anonymous reviews carry no responsibility for those views and thus there is on the face of it no power of redress - though in practice, a decision can be appealed by the author. Some argue that openness should eliminate some of the worst abuses of peer review, plagiarism and bias. The main argument against open review is that junior reviewers may be reluctant to criticise the work of senior academics, for fear of reprisals, intimidation or even the hindering of career prospects. A study undertaken by the *British Medical Journal (BMJ)* identified "a significant number" of reviewers who said that they would not review if they would be identified (thus a loss of good reviewers from the system), but also found open review had no significant effect on the review quality, the recommendation regarding publication, or the time taken to review [32]. Walsh suggests that as most research has been conducted on medical journals, similar studies should be carried out in other areas to determine whether these results are generalisable [33].

Once the reviewing practice has been decided, the editor selects independent reviewers (fellow researchers, academics or subject specialists) to send submitted papers to. Sometimes they are members of the Editorial Advisory or Review Board, or the editor's colleagues from his or her personal network (if more suited to the subject area under review). In double blind reviews, all information identifying the author is removed and the paper is coded before sending to the chosen reviewers. Finally after review, the paper is rejected, accepted or returned for revision. Reviewers' comments are sent to authors, including suggested amendments, which if made

would allow the article to be accepted. Articles may go through more than one set of revisions. It is widely agreed that this improving function by referees is of value in maintaining the overall quality of scholarly literature. Lock points out that as many as 80% of published papers receive some revision [11].

4.2 Has electronic publishing changed things?

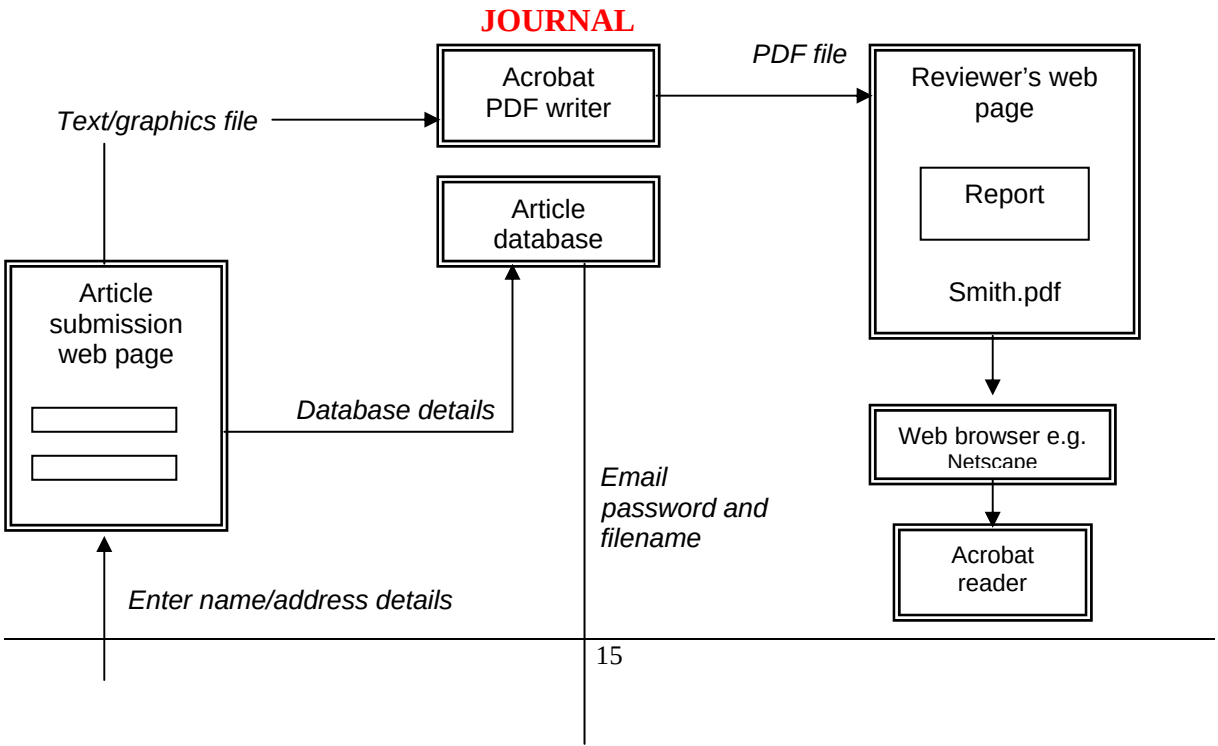
With the growth of electronic media, it now possible for material to be freely available via the Internet and for readers to comment on archived articles – a way to engage the community in the review process itself. A full discussion of this issue can be found in [31, 34, 35]. The main flaw of post-hoc review and commentary is that peer-review is meant to be a filter, saving researchers with finite time and resources, the risk of reading and using unreliable work. It needs to be done by qualified experts in advance, not afterward, by self-appointed commentators.

Harnad takes this debate further in [36, 37, 38]. He believes that the results of scholarly research should be freely available to all both before and after refereeing, i.e., that academics should make their research papers/articles freely available electronically in an un-refereed form prior to publishing (when the refereed paper is finally accepted for publication, the author can replace the preprint with the refereed version). He argues strongly for the advantages that open access both before and after peer review brings and doubts that "journal editors and referees (who, after all are the same community as authors), will long collaborate with policies that are no longer either justified or necessary, being now so clearly designed solely in the interest of protecting current 'subscription, licence, pay-per-view' revenue streams, rather than in the interest of disseminating research" [39]. He is an advocate of open access to both embryological stages of a work, but **especially** the final peer-reviewed draft. He is a strong advocate of classical peer review, until and unless an alternative is found and tested and

demonstrated to yield at least comparable quality. Open access should not be confused with open review! *However*,but argues that the increased open scrutiny that open access provides to the unrefereed draft can also help improve it and catch errors or omissions.

There are many possible ways of undertaking electronic submission and review. The method adopted by the ESPERE Project (Electronic Submission and Peer Review) [40] is shown in Figure 2 below as a means of illustrating the changes from the traditional model shown earlier in Figure 1. ESPERE began as an Electronic Libraries (Elib) project, to investigate the technical and cultural issues of peer review in an electronic environment. Since 1998, it has been funded directly by the publishers involved [(6)]. In May 2000, ESPERE launched its own peer review software and the first journal to use it was *Proceedings: Biological Science* published by the Royal Society. Now more than 20 journals published by six organisations use the system (with over 3000 submissions to date) and since 2003, all submissions for other Royal Society journals use the same system.

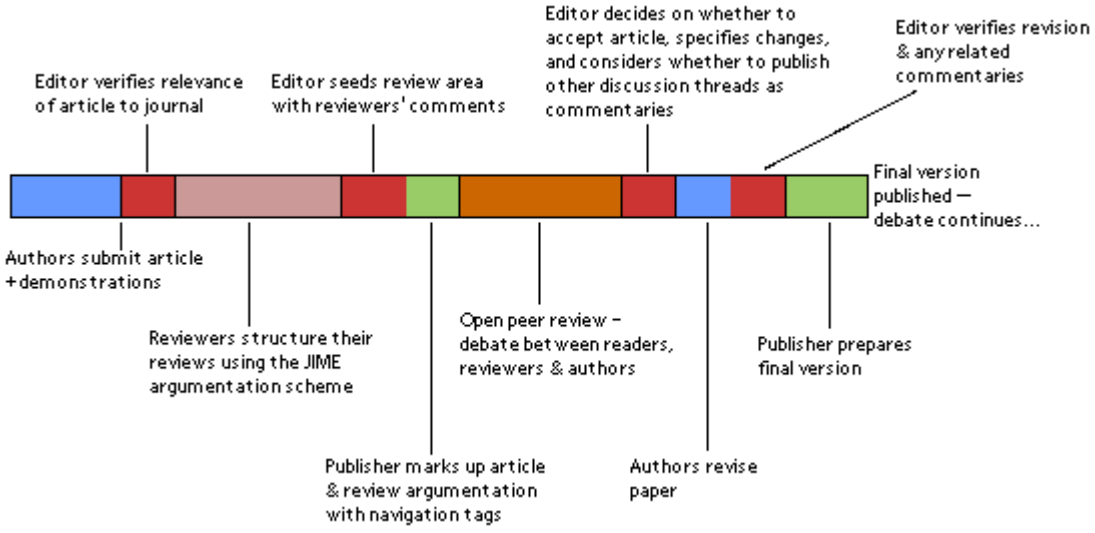
Fig. 2. One method of web based submission and review.



Source: [41].

Another model is being developed by a new scholarly journal, the *Journal for Interactive Media in Education (JIME)*, and is shown below in Figure 3 [42]. Utilising Computer-Supported Collaborative Argumentation (CSCA) technology, reviewers return their comments to the editor in the *JIME* argumentative format. The editor pulls together all the reviews to seed the argumentative debate. The article under review and the reviewers' initial comments are then published on the Web, and the review process moves into a phase of open peer review, in which authors, reviewers and readers can engage in debate. The editor then decides whether the article should be accepted, and formulates change requirements for the authors. It is also possible to publish resulting discussion threads arising during the review process that could be distilled into commentaries for publication with the final article. These experiments are interesting, but it is not at all clear whether they will scale up: referees are a rare, overused resource. A referee reluctantly referees at the call of a reputable editor of a reputable journal.

Fig. 3. Lifecycle of a *JIME* article under review.



Maintaining the quality of academic publishing is accepted as essential, but as yet no mechanism other than peer review has been suggested for achieving that quality. However, the

way in which it is funded and undertaken may be changing. In 1997, Wood looked at costs associated with electronic peer review (for the *ESPERE* project) and found that there is potential for reducing associated costs but that problems remain over handling complex formats and over lack of (at the time) standardisation (graphics, software, email) [43].

Some publishers have embraced the new technology, but few have changed the fundamental peer review process, e.g.,

- Emerald's website (management and library and information journal publishers) hosts the Literati Club for authors [44]. Giving specific information on the peer review process, it has been developed from PeerNet, an electronic double-blind reviewing system specifically for Emerald titles.
- Elsevier has instituted the Author Gateway on its website. This provides information for authors before and during the submission process and enables tracking of an article's progress afterwards [45].
- The Los Alamos pre-print server [5] created by Ginsparg in 1991 allows the physics community to exchange un-reviewed preprints and refereed postprints electronically. It is now their main method of communication, allowing the rapid distribution of research results.
- Biomed Central (BMC) has adopted a business model allowing unlimited free access for users to its journals with the cost of editorial work, peer review and publication being met through a charge to the author or author's institution. This (European) model has been adopted by the Institute of Physics for a new journal under SPARC [46], which began publication in May 2000. All manuscripts are peer reviewed via the Internet, so that the time between submission and publication is only a few weeks.
- The e-journal *Psycoloquy* [47] began publication in 1990. Edited by Stevan Harnad, it is described as an international, interdisciplinary e-journal of open peer commentary, hoping to serve as a model for electronic scholarly periodicals. It is described in [48]. It is classically peer-reviewed, and differs in that it offers open peer commentary **after** an article has been accepted for publication. The same principle was used for 25 years with *Behavioral Brain Sciences*, a paper journal, but there is no doubt that the electronic medium speeds up and simplifies this process.
- Many other journals have explicit procedures for online peer review include *First Monday* [49], *Journal of Artificial Intelligence Research – JAIR* [50], *Sociological Research Online* [51] and *Internet Archaeology* [52].

- In a move away from anonymity for referees, in 1999 the *BMJ* changed from a system of closed review to one of open, signed review on submitted papers. It intends to go further, listing reviewers at the end of published articles [53] and allowing authors and readers to observe the peer review system on the Internet, contributing comments. The *BMJ* sees such transparency as a way of showing that journal peer review does add value to the scientific process and has a place in an electronic environment. It is not yet clear whether this experiment is succeeding .

Harnad sums up the impact of new technology, stating, "refereed journal literature needs to be freed from both paper and its associated production costs, but not from the process of peer review, whose "invisible hand" is what maintains its quality" [55].

5. Quality and research publications

5.1 Assessing journal quality

For many years now the RAE has involved a process of assessing published output, and arguably this is the single most important component of the RAE score a Department is awarded. But how do the assessors judge the published output? Many questions surround the issue of quality as applied to research publications. The concept of quality of published output is very subjective and is not explored in depth in this article. What is a high quality article to one person is poor quality to another. Boaden and Cilliers propose that research is both a product (providing publications, trained researchers, etc) and a service (problem solving, advancing knowledge, etc) [56]. They discuss theoretical frameworks for measuring such performance, concluding that quality is one aspect only. They recommend a multi-factor approach be adopted to take account of other factors such as dependability, flexibility and innovation – the latter being especially important, as it has already been seen to suffer under the traditional peer review process (see Table 2).

The quality of research is by definition hard to evaluate using measures such as numbers. What is clear is that reputation signals play an important role in this evaluation process. Some of these are tacit, most subjective and none (within the RAE process itself) made explicit. Publication outlets themselves therefore become a comparative measure of the research being published. Journals could be said to represent a preference hierarchy, at times quantified by ranking measures or impact factors; individual researchers somehow prioritise journals relevant to the field in which they would like to see their work published, sometimes by perceived reputation alone. These reputation signals may contribute to a type of 'halo effect' [57], whereby departments could benefit unduly from the reputation of the institution as a whole, where the messages received by the reading audience may connect a particular piece of research with a particular university, or an article benefit either by the reputation of the publication vehicle (i.e., the journal) or by the reputation of the author's name or institution. Ali *et al.* found that a journal's reputation is partly founded on the reputation of its editors and referees [58] and Wells concludes that the outlet also affects the perceived quality of the research itself [59].

There is a wide range of literature relating to journal quality evaluation. However, with no agreed framework and large differences in the methodologies and constructs behind evaluation (many subject or discipline specific), it is very difficult to compare journal quality. Constructs such as quality, impact, importance and prestige can be and often are used interchangeably to create various hierarchies of journals. Hirst points out that "this lack of conceptual development has made it difficult to compare the ideas and results of previous research studies" [60] and suggests that because journal quality is multi-dimensional, research using single item scales to measure perceptions has not been successful in determining how academics rate journals. Hirst developed a measurement of 'journal research standing' to determine a journal's status within a discipline, taking into account these multi-dimensional factors.

Ali *et al*'s study developed checklists for assessing the 'quality of research articles' and the 'quality of journals' [58]. Although intended for use in tenure and promotion decisions, their method for assessing journal quality uses a list of 16 weighted variables (e.g., editorial reputation, impact factor and circulation) and serves to illustrate the complexity of such decisions.

In the UK, after the early RAEs moved from quantitative to quality assessment, it became apparent that benefit could be gained by having publications in the most prestigious journals. Some academics began investigating the journals themselves, developing ranking studies in order to generate 'lists' of core journals designed to help themselves and sometimes the wider community to target their output. Many disciplines began to assess the quality of their relevant journals. Some examples are: Accounting [61]; Law [62]; Information Science [63]; Business and Management Studies [64]; Library & Information Science [65]; and History [66].

5.2 Citation and journal Impact Factor (IF) Analysis

Two dominant methodologies are used to quantify the quality of published research; studies on peer review and citation analysis (including IF). Peer review studies (in this context) often take snapshots of academic opinion regarding various journals (almost always subject specific) and citation analysis measures the extent to which articles are cited (or referenced) by other published material. In some instances, these are assessments by and of the research outlets themselves and not the research. Comparisons are made all the more difficult by the fact that the nature and quality of individual articles within journals will vary, just as the nature of individual journals varies. With no objective process, system or measurement in view, individuals and departments are left with subjectivity only. This has led Campbell *et al.* to state "it is *perceptions* of quality that influence departmental decision-making in connection with the RAE" [67]. Peer-reviewers "perceive" quality (and editors count their votes); reader/authors "perceive" quality (and scientometricians count their citations). Assessors count performance

indicators and make perceptual judgments too (possibly allowing the numbers to speak entirely for themselves: that's a perceptual judgment too.)

Wouters links [68] together peer review cycles (Figure 4) and citation cycles (Figure 5), to give a model, which allows an understanding of the various feedback processes between the two (Figure 6).

Fig. 4. Peer review cycle

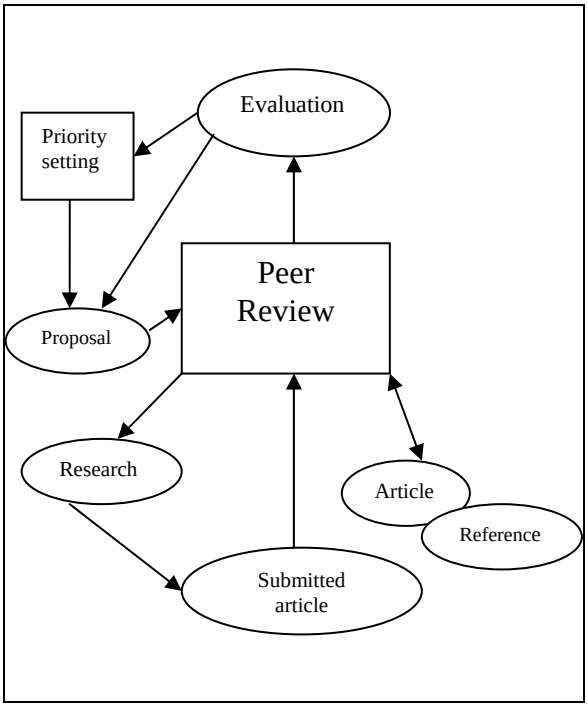


Fig. 5. Citation cycle

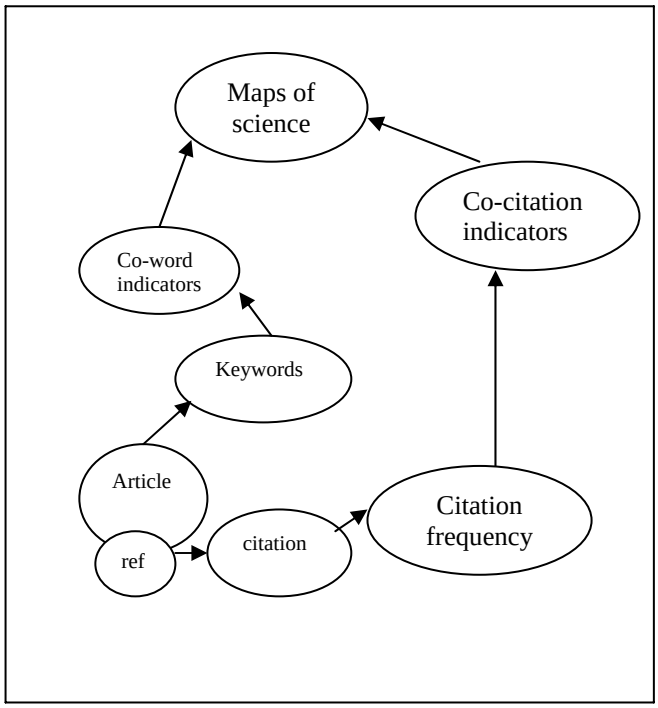
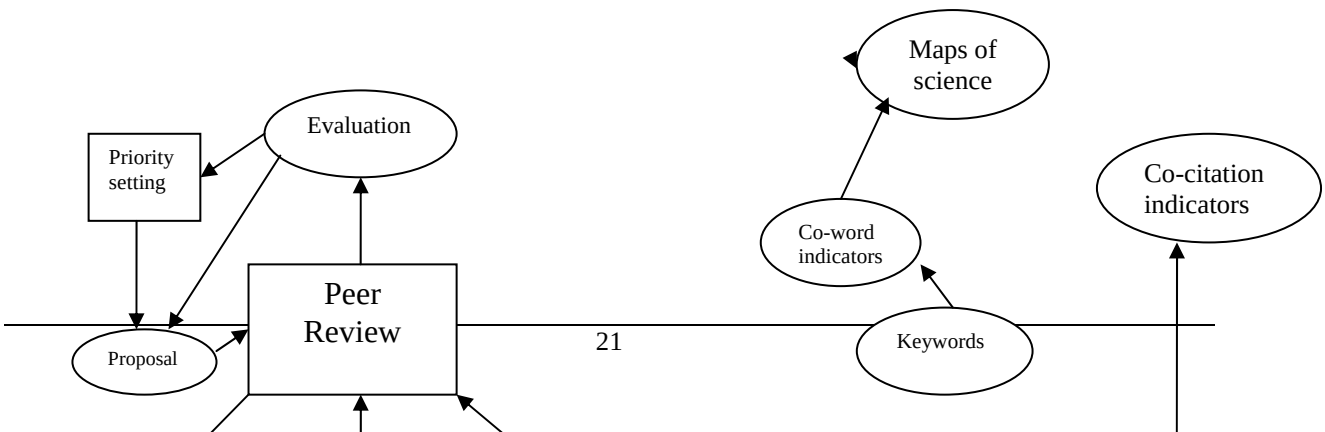
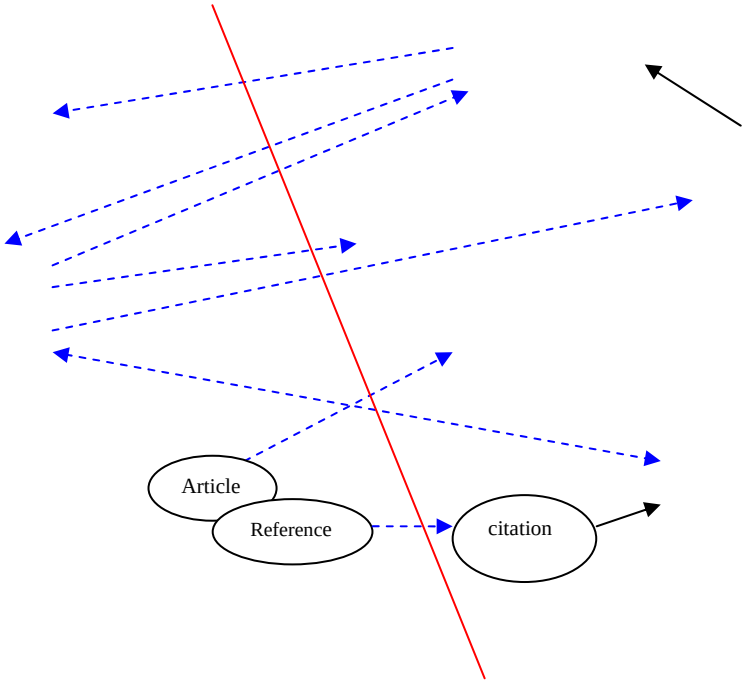


Fig. 6. Interactions of peer review cycle and citation cycle





Wouters' proposes that if science is an information processing cycle, its quality is maintained by the peer review cycle, based on the intellectual content of the literature (itself a specific representation of the research results). The second cycle (citation) is based on formal information about this, neglecting the intellectual content of the scientific literature. Thus the combination of these models enables a new view to be taken of the interaction between scientific knowledge production and evaluation.

The right hand side (right of the sloping line) in Figure 6 represents the domain of the citation cycle with its formal representation of scientific literature. The left hand side represents the domain of peer review procedures, stressing the cognitive dimension of science. Wouters claims that interactions (if realised) may influence the scientific system both at the level of the individual scientist and at the level of science policy. The danger here could be that evaluation would get two different types of input; one representing the conclusions of field specific experts (peer review) and the other representing scientometric expertise (citations). Wouters claims that, because of this, "the field specific scientist no longer has the monopoly position in evaluating science." [69]

Many calculations of research performance are based on such bibliometric measures, analysing publication references and their citation frequency. As pointed out by Baird and Oppenheim [70], citations are a well-established measure of esteem or impact, but are only a guide and partial measure of research performance. Even so, citation use is a well-known and accepted practice among academic authors. Valauskas states, "a given idea in a paper is legitimised by its publication" [71]. Citations quantify references to such ideas, but can only partially reflect quality.

To a large extent, researchers judge for themselves which journals are the most important in their field, but new researchers and junior academics need help in identifying these journals. Citation analysis assumes that references to a particular journal article reflect the scholarly impact of that article on the author of the citing work, and that this also reflects the impact of that author's work on scholarship and research in the field. At the very least, a citation implies that the cited work has been **used** to the extent (one hopes) of being read and then cited in a later work. This is by way of contrast of an **uncited** work, which may not have been read at all, and certainly has not been discussed in a later work. In addition, it may also be said that the accumulated citations to all articles published by a particular journal is an indicator of the impact of that journal on the relevant discipline. This commonly used measure is the journal Impact Factor (IF). Citation data and journal IFs can be accessed and compared through the *Journal Citation Reports* databases (hereafter *JCR*) published by the Institute for Scientific Information (ISI) [72]. Since the 1960s, ISI has scanned references in major journals and collated citations to previously published work. *JCR* is compiled annually from the complete ISI databases, comprising the *Science Citation Index (SCI), Arts and Humanities Citation Index* and the *Social Science Citation Index (SSCI).* Within *JCR*, citing journals are listed by subject category (sometimes more than one) and there is also an alphabetical listing of journals cross-

referenced to the subject categories. It is therefore possible to determine a journal's position within a subject category by IF.

References in the ISI database show how many times articles have been cited within the previous two-year period, and by whom. From this, the annual citation rate of papers by an author or research group can also be calculated. However, *JCR* should not be the sole source of information used when comparing and evaluating journals. Quantitative citation data are intended to complement NOT replace traditional qualitative and subjective inputs, since the ISI database is not without its problems; for example, in the past secondary authors were not included and some publication formats are cited more than others. Although there are disadvantages in the use of IF (Impact factor) as an index of research excellence, particularly across different specialities, it is probable that some departments already use IF measures to assist their submission decisions for the RAE. As a consequence, many researchers find themselves under pressure from heads of departments to send their papers to journals with the highest IF, which is not necessarily the best strategy.

There is a large body of literature investigating the pros and cons of citation analysis and IF. Some limitations of use and of citations more generally, are summarised in Table 5.

Table 5.
Problems associated with the use of citations and journal IF.

• Citation frequency may not be a valid indicator of scientific quality.
• Citing an article is not confirmation that the article was read or even understood.
• All citations are equal and can include 'negative' citations. One of the most cited scientific papers of the past ten years was the false alarm about low-temperature nuclear fusion, hundreds of papers have cited it: along the lines of "despite this report there is in fact no evidence for cold fusion". Brown states that this practice gives citations (as a measurement of quality) "a spurious air of relevance". [73]
• Inaccurate citations. The output data is at the mercy of the input data, which undoubtedly contains mis-spelt names and typing errors in the list of citations, which may then be quoted by other authors.
• The IF for journals from a number of countries, including Russia and China, is reduced by a publication lag, compounded by delay caused by translation before the paper is entered into the database.
• IF is heavily influenced by a form of self-citation [(7)] (articles tend to preferentially cite other

articles in the same journals) and the national bias of North American scientists to cite each other.
• Review articles are heavily cited and inflate the IF of journals.
• Citations to any type of article (including letters, editorials, reviews, communications, meeting abstracts) are used to calculate total citations, but this is then divided by the number of normal articles only. As a result the IF can be greatly inflated relative to journals that lack such items.
• IF is a measure of average citation impact, not individual citation impact, so an IF cannot be used to measure individual performance.
• The average IF of the papers of an individual scientist may not agree with peer assessment.
• The journal IF is not valid for the assessment of the quality of individual articles or authors.
• The database (ISI) has an English language bias, being dominated by American publications.
• Confusion may arise over incorrect citations i.e., authors with the same name within or between disciplines.
• Long articles collect many citations and give higher IF.
• Small research fields tend to lack high impact journals.
• New journals fare badly, due to the two-year time lag.

Source: Adapted from [74] and [75].

Reviews of citation studies can be found in [76] and [77]; citation analysis and electronic publication is assessed in [78]; citations and the RAE examined in [79, 80 and 81]. In this last article, Holmes and Oppenheim point out that, despite the many criticisms of citation counting, citation studies consistently prove to be statistically significant when correlated with other measures of eminence. In an overview of research evaluation, Hemlin also confirms a strong positive correlation between citation counts and other means of measuring research excellence, despite concerns over the validity of citation counting [82].

On the other hand, in an exploration of why journal IF should not be used for evaluating research, Seglen concludes "the journal cannot in any way be taken as representative of the article, even if it could, the IF would still be far from being a quality indicator. Citation impact is primarily a measure of scientific utility rather than scientific quality and authors' selection of references is subject to strong biases unrelated to quality" [83]. An alternative, not considered much in the literature, is the evaluation of author IF or article IF. It is surprising that so many senior managers in academic institutions pressurise their staff to get published in a journal with a high IF, rather than concerning themselves about the IF of their individual members of staff.

MacRoberts & MacRoberts summarise the many (practical) problems associated with citation analysis [84]. Errors include mis-spelt author's names, omission of author's middle initial, punctuation errors, erroneous article/journal titles, missing or added words, errors in pagination, volume numbers and dates and inconsistent journal abbreviations. Both Sweetland [85] and Pope [86] have studied citation accuracy. Sweetland's study looks at citations that could not be located and found error rates ranging from 11% - 54% depending on the discipline. The most common error was in author name and article and/or journal title. Pope's study divides errors into two categories, major errors that hinder article lactation and minor errors that do not, but are still wrong. Pope concluded that library studies, which might be expected to be a leader in citation concerns, had a high error rate in the ten journals studied, of approx 33%. Despite this "noise", citation counts strongly correlate with all types of evaluations of quality.

Citation accuracy is vital if the trail of scholarly evidence is to be preserved and presented in an accurate and useable manner. Responsibility for accurate citation lies with the author making the citation, not with referees or editorial staff. Incorrect citation also hinders librarians (since if citations are incorrect or unverified, requests for inter-library loans cannot be fulfilled) and administrators (in referencing for submission to assessment exercises such as the RAE). We believe that the online medium will play a strong role in correcting and regularising citations; this is another factor in favour of open access provision.

Citations save individual readers the work of filtering out insignificant texts [87]. Kochen notes that another purpose of citation is to acknowledge the 'intellectual debt' [88] owed to previously published research, publicly recorded in the references listed by authors, and as such remains an important indicator of how current researchers are using previous journal literature. Cronin confirms that editors and advisory boards feel that referencing is one way in which the academic community can distribute proper recognition [89] and also states that some editors and advisory boards feel that authors often fail (intentionally or not) to cite all pertinent work.

Brown [73] argues that the citation rating has become a victim of Goodhart's law [(8)], which says "any performance indicator loses its value when it becomes a policy target" [90]. Once a performance indicator (PI) is defined, those whose performance is being evaluated will seek out ways to improve their position as measured by the indicator (in the instance of the RAE, researchers will target the best ways to increase or improve their publication record). Strathern restated this law more succinctly as "when a measure becomes a target, it ceases to be a good measure" [91].

Very few studies comparing RAE scores with citation counts have been carried out; where they have been undertaken, there are statistically significant correlations between assessments made by RAE panels and citation counts, see, e.g., [80, 92,93].

6. Peer review and the RAE

It will be clear from the above arguments that peer review, whilst attracting some controversy, is generally accepted as the best method available for evaluating draft research output before it gets published. The RAE adds another level to this by making that refereed output subject to another round of review. Initially, the original quantitative approach of assessing research relied on the extent to which research was published, i.e., publication counts, as the basis for developing a metric of the research in question, and included numeric totals of publications per head. The move to qualitative measures meant a greater reliance on the system of peer review. From 1996, the RAE "secondary peer review" was taking peer-reviewed articles and other work and analysing them, whereby, "the 2001 RAE will follow broadly the same approach as previous exercises. Submissions will be made to a number of subject-based Units of Assessment (UoAs) and information supplied by Higher Education Institutions (HEIs) will provide the basis for *peer review assessment* of research quality by specialist panels" [94]. The

importance of peer judgement as an evaluation process for the RAE (as distinct from formulae or quantitative assessment) had now been underlined by the Funding Councils.

By definition, it is a process that depends ultimately on professional judgements that are not transparent. Subject panels within the RAE are selected by the Funding Councils following nominations from the academic, industrial and practitioner communities and they evaluate all submissions for assessment within that UoA. From 1996, panel criteria identified the quality of the research-active staff publications submitted by institutions as their most important criterion, but added that "they should be considered within the context of the broader picture of each department's research activity which the panel will build up from the full set of evidence available to it" [95].

However, this was not universally popular. Willmott argues, "the RAEs have evolved as a potent mechanism for justifying the withdrawal of public research funding from an increasing proportion of academics and departments. The exercises, legitimised by peer review, have facilitated a simultaneous expansion of teaching and research activity in higher education and a reduction in unit costs demanded by business leaders". This legitimising of the process by peer review may be a double-edged sword, as Willmott concludes "is peer review a plausible way of characterising this process, given that panel members are not chosen by their peers and that neither panel members nor the wider constituency of researchers play a part in the overall design of the exercise?" [96].

Parallel to the case of journal reputation, Bourke has stated that panels undertaking their review of research publications to an extent rely "heavily on surrogates, of which the most important is the evaluation processes of scholarly journals and book publishers" [97]. Willmott extends this to the RAE itself when he says, "RAE performance indicators produce an important halo effect that exerts a hegemonic influence in defining where good research is being undertaken" [98].

Just as one of the alleged disadvantages of journal peer review is one of bias (see Table 2), Doyle *et al's* statistical work on the 1992 RAE led to similar accusations of the exercises up to 1996, and specifically (for UoA 43) that "in its implicit policies the panel has fallen prey to just about all the different variants of the home team bias. English universities are favoured, old established universities are favoured, panellist's own universities are favoured and British journals are favoured, although marginally so. Large institutions are also favoured" [99]. Such judgements led to major changes for more recent exercises, in terms of the need for assessors outside the university sector and outside of the UK and for panels to consider rankings in the absence of affiliated panel members. In endorsing more 'open' review, Walsh *et al*. concluded that increased accountability in the system is essential [100]. They also recommend further research of the peer review process, not only within journals but – as similar arguments could be directed at the process by which research funding is awarded – also by the research funding bodies.

Cost has also been a criticism of the RAE, with Cooper and Otley suggesting that "if costs were reduced, it is likely that the peer review of research quality (i.e., the panels) would be lost and cruder assessments of journal quality and citation analyses used" [101].

The current RAE grading system is based on research being judged as either of 'international' or 'national excellence'. Although for 2001 the international aspect of research was confirmed by overseas panel advisors (in effect yet another layer of peer review), no clear definition [(9)] of what would constitute 'international excellence' was given. Instead, each panel had an element of freedom in deciding what would constitute the term [103]. HEFCE stated, " 'national' and 'international' refers to standards, not to the nature or scope of particular disciplines"[102]; this meant that research would need to be recognised and respected abroad rather than just published in an international journal.

These problems and issues around citation analysis, IF and peer review, have very real effects at a micro level, i.e., on individual journals and on subject specific assessment within UoAs, as illustrated by the following examples: -

- In an Editorial for *Plant Pathology* (a UK journal) Brown pointed out that this journal had the second highest citation rating behind *Phytopathology* (a US journal). This (he says) would seem to be because the scientific community in the US is larger and therefore cites more frequently, thus "the difference in citation ratings has had the absurd effect that some research directors, both in the UK and Europe, have all but obliged their staff to publish work on diseases which are most significant in Europe, in an American journal" [73].
- Similar problems over peer review have arisen for specific subjects and departments as a result of RAE assessment, e.g., within Philosophy (UoA 62). In a critique of the 1996 assessment, Sayers asks, "Who are my peers?" and states, "the panel is a quango, with all the secretive and undemocratic features typical of such bodies. How its members are chosen is a mystery. Little attempt is made to present them as representative of the different schools [of thought] and approaches in the field" [104]. He questions what actually constitutes a peer group in a subject like philosophy, and calls for selection of the panel by an open process, thus able to represent the diversity of approaches within the subject.
- Following the 2001 RAE, the panel's performance within UoA 21 Environmental Science was criticised by Watts, who identified this UoA as the weakest subject area, with the average grade below the threshold for government funding [105]. Watts' outlines statistical work examining predicted grades against actual grades, and showed that accuracy of predictions for UoA 21 and other UoAs were different with a confidence of 99.5%. Watts investigated various possibilities for this result, including importantly panel performance, asking if it could be that the members of UoA 21 panel were not representative of those they were peer reviewing? He questioned the expertise of panel members in the subject area being assessed and their 'ownership' of it (the majority of panel members were in fact earth scientists). Without government funding, some departments may have to close down, as they are in the position of having 'failed' RAE assessment after major investment and find it impossible to find out exactly why this is. He pointed out that there is still no appeal process against gradings, unless the results of panel assessment were to be overturned in court. Publication of the 2001 RAE results was followed by a memorandum from the Institute of Environmental Sciences to

the Select Committee of Science and Technology Evidence in January 2002, which stated that "the outcome of the UoA 21 panel deliberations appears to be significantly different from the RAE as a whole, from UoA 20 (joint panel with Earth Sciences) and the umbrella panel of sciences. There are only two 5* (both of whom had panel members) and two 5 universities equating to just 12% of the submitted total. This is the lowest % of 5* departments in all of the panels" [106]. The joint UoA 20 and 21 panel contained just two members whose institution submitted to UoA 21, thus (according to Watts) calling into question the peer review capacity of the panel. He concludes, "however the peer review system for the next RAE works, peer reviewers must be representative peers of those they are reviewing" [105].

It seems that the Funding Councils have more work to do in persuading many within the community that the RAE works as a system of review by peers. Nash states the 'rules' of recognition of one's peers as being "clearly those selected for the purpose must be capable of critically examining a piece of research in their chosen field, which of course necessitates that they be able to understand it in its own terms" [9].

Rules imposed by the RAE have led to the 'measurement' of quality, in a large part by the role ascribed to journals, in the process itself. In the last two RAEs, academic journals increasingly dominate the research quality scoring system (see Table 6). We believe that in many cases, the RAE has led authors and/or senior management in Universities to view successful publication in a high quality journal itself as the objective of their work, rather than the focus being on the dissemination of knowledge and ideas. No doubt publication in a high quality journal has always been the aim of most researchers, but we believe the RAE has increased this pressure. Consequently, if a performance measurement system focuses upon a particular mode of knowledge dissemination such as 'journal quality', then as Parker *et al*. indicate, a department with research ambitions is likely to adopt that system's definition of acceptable scholarship [107] and referring back to Goodhart's law, it thus may cease to become a good measure.

Table 6.

Number of journal publications submitted to RAEs 1996 and 2001

1996 RAE			2001 RAE		
Output type	**No. of items**	**% of total submissions**	**Output type**	**No. of items**	**% of total submissions**
C. Journal articles	**132,077**	**62%**	**D. Journal articles**	**143,362**	**69.7%**

Source: [44].

As regards reputation, Talib [108] points out "one cannot ignore the fact that the reputation of the journal is likely to be taken into consideration when quality assessment is undertaken by the [RAE] panel. It is speculated (and imperative) that assessment panels do regard the journal review process as a preliminary quality evaluation". This is especially relevant as he goes on to note "that anecdotal evidence suggests that RAE panel members don't read all papers submitted to them, and at times rely on where they are published" [108]. (Indeed, there is more than anecdotal evidence that this is the case.) Writing about the 1996 RAE in Law, Barnard suggested, "as for relying on the prestige of the journal in which an article has been published, the market for law review articles is so replete with imperfections that this approach is inappropriate for any serious effort at quality assessment" [109]. The Library and Information Management's panel made similar comments. By 2001, there was still no consensus on this issue. The Working Criteria of at least two UoA panels illustrate differing views, i.e., Clinical Dentistry (UoA 4) "No regard will be given to the standing of the author(s) in the field or, in the case of journal articles, the journal in which the work appears." In contrast, Statistics and Operational Research (UoA 24) stated that "for all cited works, whether read or not, the Panel will take note of the perceived editorial standards of journals" [110].

Given the very nature of peer review and the (as yet) unmeasured influence of perception and reputation, it must be debateable just how far the subjective assessments in the RAE can provide a reliable 'measure' of quality. Indeed, it is likely that sometimes members of the RAE review panel are the same individuals that refereed an earlier version of the articles being

considered. Back, referring to anonymous review (both of journal papers and RAE assessment) gave voice to such concerns when he asked, "can we continue to defend a situation where the fruits of our intellectual labours are decided by nameless judges, who are not held accountable for the content of their opinions?" [111].

Gillette's work, mentioned in the Introduction, on peer review as a research performance indicator [(10)] concludes that only journal peer review can be a true performance indicator (as opposed to grant giver and impressionistic peer review), in the sense that it relates outputs to inputs. He states that "indices based on journal peer review are able to supply reasonably valid measures of departmental performance, provided that all research output, by every academic is examined (not just a sample of the department's best work)" [113]. Although written in 1989, this would seem to detract from HEFCE's use of selective qualitative assessment as the best way of utilising journal peer review as a research performance indicator.

Some of the team working on the *Robert's Review of Research Assessment 2003*, believed that it could be time for a move (back?) to assess research via PIs, but came to recognise that whilst metrics may be useful in helping assessors to reach judgements on the value of research "we are now convinced that the only system which will enjoy both the confidence and the consent of the academic community is one based ultimately upon expert review. We are also convinced that only a system based upon expert judgement is **sufficiently resistant to unintended behavioural consequences to prevent distorting the very nature of research activity**" [114]. The report also suggested that any PIs used in the next RAE should be discipline specific. Many organisations submitted background evidence to the *Robert's Review of Research Assessment* and one study [115], found that academics surveyed favoured the retention of expert panels, but with improvements, such as: -

- clear rules and transparent procedures, especially in who was appointed to the panels, and in the criteria used for assessment
- should still involve peer recognition

- be appropriate to each discipline
- be comparable

It was also suggested that panels should be 'professionalised' possibly with paid Chairs. Roberts also made a point regarding the use of proxies (such as presumably output vehicles) "it is generally accepted that proxies can properly be used by reviewers to help them arrive at judgements about the quality of the research" [116]. This last point needs to be clarified.

7. Conclusion

There is much work still to be done in this area before the next RAE. One important proposal is for RAE CVs to be linked electronically to university eprint archives [117]. This would allow research to be continuously assessed in a more cost effective way, as well as allowing free access to all research publications, without the 'access-toll barriers' of subscriptions – thus increasing their use and impact. This is confirmed by Lawrence, who states that "articles freely available online are more highly cited" [118].

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. Henkle notes that the RAE "captures department reputations and sets them out in public form…This exposure has left some (institutions and individuals) winners and some losers – whilst others are still learning to play the game!" [119]. Piercy too alludes to 'the game' – "the game is that we are judged primarily by other academics, on the basis of publications read only by other academics and research grants awarded by academics to academics!" [120]. Indeed, the academics doing the judging 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 don't. We conclude that because of the many criticism 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.

Acknowledgements

We are grateful to the anonymous referees and especially to Dr Stevan Harnad for their helpful comments on an earlier draft of this paper.

REFERENCES

[1] R. Gillette, Research performance indicators based on peer review: A critical analysis, *Higher Education Quarterly* 43 (1989) 20-38.

[2] *PostNote No. 182 on Peer Review,* (Parliamentary Office of Science and Technology, Millbank, September 2002). Available at: www.parliament.uk/post/pn182.pdf *(Post is an office of both Houses of Parliament charged with providing independent and balanced analysis of public policy issues.)*

[3] J. Ziman, Public *Knowledge* (Cambridge University Press, Cambridge, 1968).

[4] J. Willis, *Bridging the gap between traditional and electronic scholarly publishing,* EDUCOM Conference: Portland, Oregon, 1995. Available at: www.coe.uh.edu/~brobin/Educom95/EducomJW/index.html

[5] *Los Alamos* pre-print server website. Available at: http://arxiv.org

[6] J. Smith, Peer-review – a vital ingredient, *Serials*, 4(2) (1991) 9.

[7] A.C. Weller, *Editorial peer review: Its strength and weaknesses,* ASIST Monograph Series (Medford, NJ: American Society for Information Science and Technology, 2001).

[8] R.E. Abel, and L.W. Newlin (eds), *Scholarly Publishing: Books, Journals, Publishers and Libraries in the Twentieth Century* (John Wiley, New York, 2002).

[9] F. Nash, *Peer-review and the production of knowledge.* (Proceedings, Conference of the PSA Political Science Group, University of Southampton, 22 April 1996). Available at: www.soton.ac.uk/~psd/1996/peer.html

[10] D. Rennie, Guarding the guardians: a conference on editorial peer review, Editorial. *Journal of the American Medical Association*, 256 (27) (1986) 2391-92.

[11] S. A. Lock, *Difficult Balance: Editorial Peer Review in Medicine* (ISI Press, Philadelphia Pa, 1986).

[12] H-D. Daniel, *Guardians of Science: fairness and reliability of peer* review (VCH Verlagegesellschaft, Cambridge, 1993).

[13] C. Tenopir, and D.W. King, *Towards electronic journals: Realities for scientists, librarians and publishers* (SLA Publishing, Washington DC, 2000).

[14] A. Odlyzko, The rapid evolution of scholarly communication, *Learned Publishing* 15(1) (2002) 7-19.

[15] A. J. Meadows, *Communicating Research* (Academic Press, San Diego CA, 1998) 177-194.

[16] F. Rowland, Printed journals: Fit for the future? *Ariadne* 7 (1997). Available at: www.ariadne.ac.uk/issue7/fytton

[17] *Journal of the American Medical Association*, themed issue 280 (1998).

[18] V. Bence and C. Oppenheim, The role of academic journal publications in the UK Research Assessment Exercise 1986-2001, forthcoming *Learned Publishing* Jan. 2004.

[19] F. Lee, and S. Harley, Economics divided: the limitations of peer review in a paradigm-bound social science. In: D. Jary, and M. Parker, (eds.), *The new HE: Issues and directions for the post-Dearing university* (Staffs University Press, Staffs, 1998) 192.

[20] P. Evans, The peer-review process, *Literati Newsline; Special Issue for Authors and Editors*, 2 (1995).

[21] H. El-Munshid, Evaluation of peer review in biomedical publications (Editorial), *Annals of Saudi Medicine* 21 (5-6) (2001).

[22] C. Tenopir, and D.W. King, Trends in scientific scholarly journal publishing in the US, *Scholarly Publishing* 28 (1970) 135-170.

[23] B. Donovan, *The truth about peer review*. ICSU Press Workshop, Keble College Oxford, 31/3/1998-2/4/1998. Available at: www.bodley.ox.ac.uk/icsu/donovan/ppr.htm

[24] F. Rowland, The peer review process, *Learned Publishing* 15 (2002) 247–258.

[25] G. Roberts, *Review of Research Assessment: A report to the Funding Bodies*. (HEFCE Report 22, Bristol, May 2003).

[26] *Ibid, Recommendation* 1, 7.

[27] A.C. Weller, *Editorial peer review: its strength and weaknesses*, ASIST Monograph Series (Medford, NJ: American Society for Information Science and Technology, 2001) 307.

[28] D. E. Chubin, Refereeing and trilateral negotiation. In: R. Weeks and D. Kinser (eds.), *Editing the refereed scientific journal: practical, political and ethical issues* (IEEE Press, NJ, 1994) 7.

[29] *Ibid*, 8.

[30] J. Adams, et al. *Benchmarking of the international standing of research in England: A consultancy study on bibliometric analysis* (Report by Centre for Policy Studies in Education, Leeds University, 1998) 31.

[31] A. Weller, Emerging models of editorial peer-review. In: *Proceedings of the 8th International Congress on Medical Librarianship*, London, 2-5th July 2000.

[32] F. Godlee and T. Jefferson. *Peer Review in Medicine* (BMJ Books, London, 1999).

[33] E. Walsh, *et al*. Open peer review: a randomised controlled trial, *British Journal of Psychiatry* 176 (2000) 47-51.

[34] J. Peters, *An exploration of electronic peer review*, (1996). Available at: www.carbon.cudenver.ecu/~mryder/peer/peters.html

[35] R. Kling, and L. Cove, Electronic journals and legitimate media in the systems of scholarly communication, *The Information Society* 11 (4) (1995) 261-271. Available at: www.ics.uci.edu/~kling/klingej2.html

[36] S. Harnad, Scholarly skywriting and the prepublication continuum of scientific inquiry, *Psychological Science*, 1 (1990) 342-343. Available at: www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad90.skywriting.html

[37] S. Harnad, A subversive proposal. In: A. Okerson and J. O'Donnell (eds), *Scholarly journals at the crossroads: A subversive proposal for electronic publishing* (Association of Research Libraries, Washington DC, June 1995). Available at: www.arl.org/scomm/subversive/yoc.html

[38] S. Harnad, The future of scholarly skywriting. In: A. Scammell (ed.), *I in the sky: Visions of the information future* (ASLIB, London, November 1999). Available at: www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad99.aslib.html

[39] S. Harnad, Free at last: The future of peer reviewed journals, *D-Lib Magazine* 5 (12) (1999) Available at: www.dlib.org/dlib/december99/12harnad.html

[40] *ESPERE Project.* Available at: www.espere.org

[41] D. J. Wood, Peer review and the web: the implications of electronic peer review for biomedical authors, referees and learned society publishers, *Journal of Documentation* 54 (2) (1998) 173-197.

[42] S. Buckingham Shum and T. Sumner, JIME: An Interactive Journal for Interactive Media, *First Monday*

[online] 6 (2) (2001).Available at: http://firstmonday.org/issues/issue6_2/buckingham_shum/

[43] D. Wood, Electronic submission and peer review: An update on the ESPERE Project, *Learned Publishing* 10 (1997) 157-159.

[44] *Emerald publisher's website*. Available at: www.emeraldinsight.com

[45] *Elsevier publisher's website, Author Gateway.* Available at: http://authors.elsevier.com.

[46] JISC Scholarly Communications Group, *Report to RSLG*. (May 2002) 7. [Unpublished report. Available at: www.biomedcentral.com].

[47] *Psycoloquay*, e-journal. Available at: http://psycprints.ecs.soton.ac.uk/

[48] S. Harnad, Implementing peer review on the net. In: R. Peek and G. Newby (eds), *Scholarly publication: the electronic frontier* (MIT Press, Cambridge MA, 1996) 103-108.

[49] *First Monday* e-journal. Available at: www.firstmonday.org

[50] *Journal of Artificial Intelligence* e-journal. Available at: http://cs.Washington.edu/research/jair/submission-info.html

[51] *Sociological Research Online*. Available at: www.socresonline.org.uk/1/1/editors.html

[52] *Internet Archaeology* ejournal. Available at: http://intarch.ac.uk/news/rae.html

[53] R. Smith, Opening up BMJ peer review. A beginning that should lead to complete transparency -Editorial, *British Medical Journal* 318 (1999) 4-5.

[54] *Public Library of Science (PLoS).* Available at: www.publiclibraryofscience.org

[55] S. Harnad, The invisible hand of peer review, *Nature* (Nov 1998). Available at: http://helix.nature.com/webmatters/invisible/invisible.html

[56] R. J. Boaden and J. Cilliers, Quality and the Research Assessment Exercise: just one aspect of performance? *Quality Assurance in Education* 9 (1) (2001), 5-13.

[57] L. V. Jones, *et al. An assessment of research Doctorate programs in the USA (*National Academic Press, Washington DC, 1982) 5 vols.2.

[58] S. N. Ali, H. Young and N.M. Ali, Determining the quality of publications and research for tenure and promotion decisions, *Library Review* 45 (1) (1996), 42.

[59] A. Wells, *Exploring the development of the independent, electronic, scholarly journal.* [Unpublished MSc Thesis, University of Sheffield, Department of Information Studies, 1998/99].

[60] A. S. Hirst, *Measuring journal quality: developing a multi-item measure and investigating its usefulness in marketing*. [Unpublished PhD Thesis, Loughborough University, 1999, 5].

[61] T. Brinn and J. Jones, UK Accountants' perceptions of journal quality, *Accounting and Business Research* 26 (3) (1996), 265-278.

[62] K. Campbell, D. Vick and A. Murray, Journal publishing, journal reputation and the UK's RAE, *Journal of Law and Society* 26 (4) (1999), 470-501.

[63] D. T. Hawkins, Bibliometrics of electronic journals in information science, *Information Research* 7 (1) (2001).

[64] B. Pierce and G. Garven. Publishing international business research: A survey of leading journals, *Journal of International Business Studies* (Spring 1995) 69-89.

[65] S. McDonald and J. Feather, British library and information science journals: a study of quality control, *Journal of Information Science* 21 (5) (1995), 359-369.

[66] J. Morris, *The journal editor's view, History Periodicals - Past and Present Trends.* [Unpublished paper delivered at the Anglo-American Conference of Historians, Institute of Historical Research, University of London, 7 July 2000.]

[67] K. Campbell, *et al*. Journal publishing, journal reputation and the UK's Research Assessment Exercise, *Journal of Law and Society* 26 (4) (1999) 474.

[68] P. Wouters, Citation cycles and peer review cycles, *Scientometrics* 38 (1) (1997) 39-55.

[69] *Ibid.*, p. 48-49.

[70] L. Baird and C. Oppenheim, Do citations matter? *Journal of Information Science* 20 (1) (1994) 2-15.

[71] E. J. Valauskas, First Monday and the evolution of electronic journals, *Journal of Electronic Publishing* 3 (1) (1997). Available at: www.press.umich.edu/jep/03-01/FirstMonday.html

[72] *Journal Citation Reports. Social Science Citation Index*. (Institute for Scientific Information, Philadelphia, PA, 2001). Available at: www.isinet.com/isi/products/citation/jcr/jcrweb/index.html

[73] J. Brown, The citation rating game - Editorial, *British Society for Plant Pathology* Newsletter (1997) 31.

[74] Anon. Editorial, An evaluation of impact factors and league tables, *Journal of the Royal College of Physicians* 32 (3) (2002).

[75] P.O. Seglen, Why the impact factor of journals should not be used for evaluating research. *British Medical Journal* 314 (1997) 498-502.

[76] M. Liu, Progress in documenting the complexities of citation practice; a review of citation studies, *Journal of Documentation* 49 (4) (1993) 370-408.

[77] T. J. Phelan, A compendium of issues for citation analysis, *Scientometrics* 45 (1) (1999) 117-136.

[78] S. Harter, The impact of electronic journals on scholarly communication: A citation analysis, *Public Access Computer Systems Review* 7 (5) (1998) 5-34.

[79] J. Warner, A critical review of the application of citation studies to the RAE, *Journal of Information Science* 26 (6) (2000) 453-460.

[80] A. Smith and M. Eysenck, *The correlation between RAE ratings and citation counts in psychology*, (2002). Available at: http://psyserver.pc.rhbnc.ac.uk/citations.pdf

[81] A. Holmes and C. Oppenheim, Use of citation analysis to predict the outcome of the 2001 RAE for UOA 61: Library & Information Management, *Information Research* 6 (2) (2001).

[82] S. Hemlin, Research on research evaluation, *Social Epistemology* 10 (2) (1996) 209-250.

[83] P. O. Seglen, Why the impact factor of journals should not be used for evaluating research, *British Medical Journal* 314 (1997) 500.

[84] M. MacRoberts and B. MacRoberts, Problems of citation analysis: a critical review, *Journal of the American Society for Information Science* 40 (5) (1989) 342-348.

[85] J. H. Sweetland, Errors in bibliographic citation: A continuing problem, *Library Quarterly* 59 (4) (1989) 291-304.

[86] N. Pope, Accuracy of references in ten library science journals, *Research Quarterly* 32 (2) (Winter 1992) 240-243.

[87] M. O'Donnell, Electronic Journals: Scholarly Invariants in a Changing Medium, *Journal of Scholarly Publishing* 26 (1995) 184.

[88] M. Kochen, How well do we acknowledge intellectual debt? *Journal of Documentation* 3(1) (1987) 54-64.

[89] B. Cronin, Norms and functions in citation: the view of journal editors and referees in psychology, *Social Science Information Studies* 2 (1982) 71.

[90] C. Goodhart, *Money, information and uncertainty* (Macmillan, London, 1975).

[91] M. Strathern, Improving ratings: Audit in the British University system, *European Review* 5 (1997) 305-321.

[92] C. Oppenheim, Do citations count? Citation indexing and the RAE, *Serials* 9 (2) (1996) 161.

[93] C. Oppenheim, The correlation between citation counts and the 1992 RAE ratings for British research in genetics, anatomy and archaeology, *Journal of Documentation* 53 (1998) 477-487.

[94] Higher Education Funding Council for England, *Circular RAE 5/99 Assessment panels' criteria and working methods*, (HEFCE, Bristol, 1999) para 1.5.

[95] Higher Education Funding Council for England, *Circular RAE 3/95, Research Assessment Exercise: Criteria for Assessment*, (HEFCE, Bristol, 1995).

[96] H. Willmott, Commercialising higher education in the UK: the state, industry and peer review. In: *Proceedings Higher Education Close-up Conference*, University of Central Lancashire 6th July 1998, 21.

[97] P. Bourke. *Evaluating university research: The British RAE and Australian practice*. (NBEET/ARC/HEC Commissioned Report No 56, AGPS Canberra: Australia, 1997.)

[98] H. Willmott, Commercialising Higher Education in the UK: the state, industry and peer review. In: *Proceedings, Higher Education Close-up Conference*, University of Central Lancashire, 6 July 1998, 4.

[99] J. Doyle and A. Arthurs, The judge, the model of the judge and the model of the judged as judge: analysis of the UK 1992 RAE data for business and management studies, *Omega: International Journal of Management Science* 24 (1) (1996) 21.

[100] E. Walsh, *et al*. Open peer review: a randomised controlled trial, *British Journal of Psychiatry* 176 (2000) 50.

[101] C. Cooper and D. Otley. The 1996 Research Assessment Exercise for business and management, *British Journal of Management* 9 (1998) 82.

[102] Higher Education Funding Council for England, *Circular RAE 4/01, 2001 RAE: The Outcome*, (HEFCE, Bristol, December 2001) para 3.6. Available at: www.hero.ac.uk/rae/Pubs4_01/

[103] Higher Education Funding Council for England, *Circular RAE 5/99, Assessment Panels criteria and working methods*, (HEFCE, Bristol, 1999). Available at: www.hero.ac.uk/rae/Pubs/5_99/section3.htm

[104] S. Sayers, Who are my peers? The RAE in Philosophy, *Radical Philosophy* 83 (1997) 2-5.

[105] S. Watts, The RAE – Environmentally unfriendly? *THES* No 1537 (10/5/2002). Available at: www.brookes.ac.uk/schools/bms/research/ebes/watts/THESarticle.doc

[106] J. Longhurst, *Memorandum submitted by the Institution of Environmental Sciences to HM Government Select Committee on Science and Technology*, January 2002. Available at: www.publications.parliament.uk/pa/cm200102/cmselect/cmsctech/507/507ap32.htm

[107] L.D. Parker, J. Guthrie and R.H. Gray, Publish or Perish: The social construction of quality in accounting and management research. In: *Proceedings of the Fifth Interdisciplinary Perspectives on Accounting Conference*, University of Manchester, 7-9 July 1997.

[108] A. A.Talib, The RAE & publications: a view of journal editors, *Higher Education Review* 33 (1) (2000) 33.

[109] J. W. Barnard, Reflections on Britain's Research Assessment Exercise, *Journal of Legal Education* 48 (4) (1998) 490.

[110] Higher Education Funding Council for Education, *Circular RAE 5/99, Assessment Panels Criteria and Working Methods*, (HEFCE, Bristol, 1999). Available at: www.hero.ac.uk/rae/Pubs/5_99/ UoAs 4 & 24.

[111] L. Back, The devil you know, *The Guardian* (18/7/2000) 13.

[112] R. Gillette, Research performance indicators based on peer review: A critical analysis, *Higher Education Quarterly* 43 (1989) 37.

[113] *Ibid.*, 22.

[114] G. Roberts, *Review of Research Assessment: A report to the Funding Bodies*. (HEFCE Report 22, Bristol, May 2003) 7.

[115] *RAND Report*. Available at: www.rand.org/randeurope/news/rae.study.html

[116] G. Roberts, *Review of Research Assessment: A report to the Funding Bodies*. (HEFCE Report 22, Bristol, May 2003). Chapter 3 para.27.

[117] S. Harnad, L. Carr, T. Brody and C. Oppenheim**,** Mandated online RAE CVs linked to university eprint archives: Improving the UK RAE whilst making it cheaper and easier. Forthcoming in *Ariadne*. Available at: www.ecs.soton.ac.uk/~harnad/Temp/Ariadne-RAE.doc

[118] S. Lawrence, Online or invisible? *Nature* 411 (6837) (2001) 521. Available at: www.neci.nec.com/~lawrence/papers/online-nature01/

[119] M. Henkle, The modernisation of research evaluation: The case of the UK, *Higher Education* 38 (1999) 109.

[120] N. Piercy, Why it is fundamentally stupid for a business school to improve its RAE score, *European Journal of Marketing* 34 (1) (2000) 31.

FOOTNOTES

(1) It does not consider national libraries, database publishers or the role of institutional/subject-based e-print archives such as Ginsparg's [5].

(2) Research is generally supported by grants from a variety of sources, inside or outside the institution where the researcher works, and has often been already subject to peer review.

(3) Here archiving means safely storing the material, providing access methods, and creating ways for scholars to find relevant material, functions usually performed by the University Library.

(4) Journals in a variety of fields were studied and no separate costs for peer review were given.

(5) Some journals use open review, where either reviewers know the identity of the author or visa versa, but most good academic journals use blind or double-blind refereeing (where either the identity of the author or referee is undisclosed, or both), in theory ensuring impartiality. This is not always the case, since if the subject area does not support a large research community, it may be possible for reviewers to guess the origin of a paper by its content, or even the references.

(6) British Society for Immunology, Company of Biologists, Professional Engineering Publishing (IMechE), Society for Endocrinology, Society for General Microbiology, The Royal Society, University of Nottingham.

(7) The self-citation rate is a measure based on the amount of self-citations a journal receives amongst the citations of articles in that journal, or the citations of authors' to their own work.

(8) Goodhart's Law was named after Charles Goodhart a former Chief Economist at the Bank of England.

(9) The international criterion adopted equated to a level of excellence that it was reasonable to expect for the UoA, even though there may be no current examples of such a level in the UK or elsewhere. In the absence of current examples, standards were adopted from cognate research areas where international comparison does exist [102]

(10) A true performance indicator is a measure, which relates the amount of output achieved per unit of input (resources), i.e., number of publications divided by the number of research staff. Peer review is an output measure only. [112]