Consultation on the assessment and funding of higher education research post-2008

Respondent's details

Are you responding:
(Delete one)

- On behalf of an organisation

Name of responding organisation/individual

UK Computing Research Committee

Contact name

Martyn Thomas CBE FREng

Position within organisation (if applicable)

Member of the executive committee of UKCRC

Contact telephone number

01225 469441 or 07970 826812

Contact e-mail address

martyn@thomas-associates.co.uk

Consultation questions

(Boxes for responses can be expanded to the desired length.)

Consultation question 1a: Do you endorse our proposals for defining the broad group of science-based disciplines, and for dividing this into six main subject groups, in the context of our new approach to assessment and funding?

UKCRC is concerned about the haste with which it is planned to introduce this fundamental change, particularly so soon after a major review that recommended the process used in 2008. The mechanism used to allocate QR has a profound, strategic effect on university research policy, on the balance of research within and between disciplines, and on the research directions and publication policies of individual researchers. It can also have a serious impact on teaching quality and content by influencing the processes by which staff are selected and the pressures on staff to allocate their time so as to maximise research ratings. It would be incompetent and unprofessional to introduce a citation-based Research Excellence Framework until it has been established that there is an adequately complete, consistent and auditable set of data, available from multiple sources free of any commercial bias, that can be relied on to be kept up to date, that includes citations in journals, conferences, PhD theses, industrial reports and institutional repositories — and that assessments based on citation counts from these sources leads to cost-effective assessment of research quality that does not lead to undesirable changes in the way research is carried out or published or on standards or variety of teaching. We do not see any convincing evidence that these criteria have been met.

In our comments below we focus on our own discipline, although we are aware that many of the points that we make apply to other disciplines and that it may be necessary to modify the REF scheme as a whole rather than to make special provisions for computer science.

It is important to recognise that Computer Science is a very broad subject. Computer science spans research from mathematics to engineering (including aspects of the life sciences, philosophy, linguistics and many other disciplines), and the timescales over which significant influence occurs range from citations over several decades to almost immediate impact on several of the largest industrial sectors (hardware, software, new media and the rest of the digital economy).

There is a very high rate of innovation in Computer Science. One consequence is that communities develop around novel programming languages, development methods or research problems and this means that the level of citations that are stimulated by controversial proposals or fashionable ideas may be completely unrelated to their scientific value. The RAE panels have generally found it necessary to consider the quality of the journals or conferences where work is published (and often to read the papers) in order to assess research quality.

Normalising citation counts would help but, in Computer Science, the normalisation would have to be done at sub-discipline level to be meaningful, and there might be many tens of (changing) clusters that would have to be normalised differently to reach a usable result. Without such normalisation, citation counting would distort the allocation of QR between institutions and would bring about a large and undesirable change in research strategies, co-authoring and publication decisions and, over time, will have a distorting effect on areas of teaching that can be supported, including making it harder to introduce courses related to novel research topics.

It is unlikely that any system that gives the final say to numerical indices can work in other than the most homogeneous of disciplines and we therefore believe that evaluation of computer science must retain an element of human judgement that uses numerical data only as an input.

Consultation question 1b: Are there issues in relation to specific disciplines within this framework that we should consider?

Research in computer science is enormously varied. A theoretical computer scientist writes papers that could be easily compared to those of mathematicians – these papers might be very long, single authored and attract citations over a long period of time. A researcher who is active on equally important engineering topics might work with many collaborators and produce multi-authored papers mainly for conferences and the impact might be more on practitioners than on other researchers and their publications.

Anything that disturbed this balance could have a negative impact on research in computing. It could undermine research on the intellectual foundations of the subject and/or reduce its impact on a major segment of industry.

Even the notion of a "paper" varies enormously between say theoretical computing science and reporting on a large, long-term (hardware or software) engineering project. Such papers vary in length, numbers of contributors and likely citation patterns. We are therefore doubtful that they can be compared in any meaningful way. In many areas of CS research real impact comes by providing a piece of software (e.g., a compiler, an application, a simulator) or data (e.g., a file system trace) that others can use in their work. Often such things are not cited because they fall outside of journals and other publications: for example, HTML and HTTP 1.0 as defined by Tim Berners Lee, which laid the foundations of the World Wide Web, have few citations. It would be worth considering whether software download statistics are a form of citation and, if so, how they should be evaluated.

A further complication is that a successful software tool may dramatically reduce the citations to the fundamental research that it embodies, because subsequent researchers choose to cite the tool rather than the theoretical paper.

One generic issue across computing is the tendency in such a fast-moving subject to publish important results in conference proceedings. In many sub-fields, conferences are extremely competitive, heavily refereed and fully archived; in some areas, it is unusual even to follow up a conference publication with a journal submission because the delays in journals do not fit up-to-the-minute research dissemination. We believe this is different from many other scientific areas. We recommend that HEFCE commissions an analysis of the papers submitted to the 2008 RAE in Computer Science (sub-panel 23), to see the proportions contained in WoS journals, non WoS journals and conferences.

Again, anything designed purely to allocate funds that brought about significant perturbations to the norms of the subject could have a very damaging influence on research, research training and teaching in computer science .

Consultation question 2a: Do you agree that bibliometric indicators produced on the basis that we propose can provide a robust quality indicator in the context of our framework?

The UKCRC welcomes the exploration of the potential role of metrics in the assessment of research. Citation counts are clearly one candidate for inclusion in a set of metrics. However, we believe that citation counts are inadequate alone and their use might actually damage the science they are attempting to measure.

Firstly, it should be clear that exclusive (or major) reliance on any single factor is unreliable. A metric that is based on a weighted average of many correlated factors is less subject to manipulation (and any perceived manipulation can be countered by subsequent adjustment of weightings). Manipulation of relevance indices by originators of material is an all too familiar problem to those involved with web search engines.

As far as citation counting is concerned, we would find it alarming were HEFCE to base bibliometric evaluation on a *single source* of citation information (whoever the provider were to be). This would introduce a dependency and cost that would be extremely unfortunate and could introduce biases that had seriously detrimental long term effects on research and teaching, as explained above.

As far as computer science specifically is concerned, the ISI data is not fit for the stated purposes of the REF. Even the journal penetration is estimated as less than 50%. We have indicated above that it is imperative that a wide range of conferences must also both be counted as outputs and their citations credited. We also cannot see any reason for excluding references in and to books, PhD theses and institutional open document repositories.

It would also be totally counterproductive to the government's stated objectives of research impact if references in and to industrial reports/publications were excluded.

In our view, it will take many years to get an agreed source or sources of such data in computer science and to check the ability of the organisations to maintain their currency. In passing, we should also observe that the simple task of sorting out who published what has not yet been satisfactorily resolved: many of the "top rated (computer) scientists" (in citation listings) derive their placing from agglomerations of references by similarly-named, different authors. A recent paper (*URI Disambiguation in the Context of Linked Data* by Jaffri, Glaser and Millard) concluded that " ... if a person has a common name. the probability of their publications being merged with other authors will be 90%." Many publishers require that citations include initials rather than forenames of cited authors, which exacerbates this problem.

There is a similar problem in uniquely identifying a cited paper, when the citations may use different abbreviations in different publications and the paper may exist in several forms. We have seen examples where a single paper generated more than ten individual entries in WoS; considerable human effort would be required to decide which citations need to be added together.

Note that UKCRC members include internationally renowned experts in the automated collection, processing analysis and storage of information – the theories, tools and methods that underlie the proposed bibliometric indicators. Our authoritative view is that the bibliometric indicators are not currently fit for the proposed purpose.

Consultation question 2b: Are there particular issues of significance needing to be resolved that we have not highlighted?

Because of the breadth of computer science, the period over which citation counting would make sense will vary enormously between sub fields.

UKCRC considers that the comments on potential "feedback loops" lack reality. It is certainly true that an indifferent paper cannot be moved into the super-star category by "citation cliques" but it is also obvious (and born out by experience of the RAE) that whatever mechanism is chosen to determine QR funding will define a "currency" in terms of which institutions judge their staff. It will therefore be inevitable that ambitious –and busy– academic staff will look for ways to increase their citation counts and that ambitious heads of departments will reinforce these pressures. These tactical moves are likely to reduce willingness to pursue research in risky, unfashionable, and difficult, but potentially important research areas with prospects of high long term impact.

We believe that the dominant medium for dissemination and accumulation of scientific information is moving from traditional publishing towards websites. It would be short sighted not to plan for this change.

In our own subject, much effective technology transfer is by means of software tools. There must be a mechanism for assessing and recognising this factor. If this is not done, the assessment methods will motivate a move away from what is a most valuable output of researchers.

There is a potential, significant effect of the proposals on refereeing. A paper cannot be refereed without its citations, and a reviewer will usually have publications in the field that the paper is about. It is possible that reviewers will be influenced by the extent to which their own papers are cited (or that they will be suspected of being so influenced) and that this will affect both authors and reviewers. This problem is felt especially strongly in subjects such as Computer Science where the average number of citations per paper is low and the number of uncited papers is high. We are aware of some journals that have asked authors to add citations to papers in the same journal; this is likely to become more common under the proposed REF.

Consultation question 3a: What are the key issues that we should consider in developing light touch peer review for the non science-based disciplines?

As we have made clear, we do not consider that a move to a metric-based evaluation can be made in the near future for computer science. Even in the longer term, we believe that some measure of peer review must be retained.

We believe that the appropriate light touch peer review for each discipline must be developed with and agreed by representatives of that discipline.

Consultation question 3b: What are the main options for the form and conduct of this review?

For computer science, we believe that the review should be carried out by panels at the discipline level, not at the overarching level of the proposed six main subject groups, and that it should be based on peer judgement, informed by a range of metrics. The process and the metrics should be agreed with the subject panel and trialled before being finally adopted.

If disciplines are grouped together into larger units for the purpose of light touch review, then it is important that these groupings are not used for any other purpose without separate consultation.

Consultation question 4: Is there additional quantitative information that we should use in the assessment and funding framework to capture user value or the quality of applied research, or other key aspects of research excellence? Please be specific in terms of what the information is, what essential element of research it casts light on, how it may be found or collected, and where and how it might be used within the framework.

The RAE has counted memberships of programme committees, invited lectures, fellowships of learned and professional societies and other measures of peer esteem and continued activity. We recommend that the metrics to be used as input to a light touch peer review be agreed with the subject committee. If it is felt that great weight should be

placed on some formula based on metrics, then we suggest that it would be prudent to conduct the following exercise in each subject once its bag of metrics has been selected.

Experiment Gather a group of recognised experts and invite them to identify the most important papers published or research groups active in the window 20 to 15 years ago. Then ask how the formula based on the putative list of metrics would have rated the chosen papers or research groups just 5 years later.

We could not support the use of any set of metrics as the *primary* means of research quality evaluation until we have seen that they perform consistently well in this sort of evaluation.

Consultation question 5: Are our proposals for the role of expert panels workable within the framework? Are there other key issues on which we might take their advice?

Consultation question 6: Are there significant implications for the burden on the sector of implementing our new framework that we have not identified? What more can we do to minimise the burden as we introduce the new arrangements?

We have conducted an exercise on one paper that has over a hundred and fifty "references" on Google Scholar. The ISI citation count is 24; the citation count from ACM is 12 (ACM is the major professional body in international computing). *There are no citations in common between the ACM and ISI lists!* Many of the extra Google Scholar references are in industrial reports, books, PhD theses and conferences. A study based on a single paper can only be illustrative but it shows that much wider analysis is needed to find robust and valid bibliometric indicators.

We are aware that Google Scholar has some deficiencies but it exists and (at least for Computer Science) it locates so many citations that are not in the ISI data that it destroys the credibility of ISI as the basis for the REF.

The stated aims of the REF are transparency and light touch. If the exercise is to be transparent then institutions and departments must have access to the data used to generate their ratings and to the source code of any systems that create or manipulate this data and the ability to correct any errors they find. But this would create a considerable administrative burden, potentially involving checking all the institution's data held on WoS for the period of assessment. The corrections would then place a further burden on the data administrators for WoS. It would therefore be preferable to only take into account a limited number of publications per researcher. (There are other reasons to restrict the number of publications. If all publications are counted, then valuable activities such as publishing with a PhD student will be avoided for fear of reducing average citation scores. Researchers will cluster even more in popular areas of science where

citation behaviour leads to high counts, leading to less risk taking and innovation – both of which EPSRC have been trying to stimulate).

We are seriously concerned that the effort involved in submitting auditable bibliometric information (at least in computing) will take *more* effort, more often, than preparation for an RAE, without adding any value.

Consultation question 7: Do you consider that the proposals in this document are likely to have any negative impact on equal opportunities? What issues will we need to pay particular attention to?

It is recognised even by supporters of bibliometrics that the same material published (in similar outlets) under the names of, on the one hand, a widely recognised author and, on the other hand, a newcomer will receive widely differing citation patterns. Not only is this unfair on an equal opportunities front, it could have serious distorting effects on hiring patterns. Yet again, we have to point out that such changes are unlikely to be for the long term good of research in the UK.

A recent study (*Double-blind review favours increased representation of female authors*, Budden et al, Trends in Ecology and Evolution, Vol 23, No 1, currently available at http://www.onepoint.ca/Budden%20et%20al%202008.pdf) reported that there is a measurable bias in referees' reports against female authors. It would be surprising if such bias did not also extend to ethnic minorities and to authors' choice of citations. This means that bibliometrics may well have a negative effect on equal opportunities and it would take further research to discover whether the effect also exists in the current RAE and, if so, whether the REF would make it better or worse.

Consultation question 8: Do you have any other comments about our proposals, which are not covered by the above questions?

There is another factor that affects computing. As a fast moving subject, books often lag far behind research. In this environment it is not surprising that (ACM's) "Computing Surveys" is the journal with the highest impact factor, as recent surveys will often be more up to date than books. High quality survey papers are crucial for scholarship and valuable for progress, but a careless move to rewarding highly cited papers could damage research by persuading academics that it is better for their careers to write surveys than to pursue original research.

We welcome the intention to conduct a pilot study during 2008. We strongly recommend that the pilot study should be delayed until the results of RAE 2008 are known (so that there can be no question of the pilot affecting the RAE and so that the pilot can be compared with the most up to date RAE profiles). In view of the strategic importance of the REF proposals, we would also request that the pilot is conducted transparently and with independent scrutiny.