

Responses to HEFCE REF consultation questions

UKCRC REF Working Group

UKCRC

The UK Computing Research Committee (UKCRC) is an Expert Panel of the British Computer Society, the Institution of Engineering and Technology and the Council of Professors and Heads of Computing. Its members are leading computing researchers from UK academia and industry. Our evidence reflects the experience of researchers who each have an established international reputation in computing and many members of the 2008 RAE CS&I sub-panel are also members of UKCRC.

We would recommend that the time-table for the REF in 2013 should be delayed by 12 months. This would give more time for consultation on the final proposals which are likely to result in some reorganisation of panels and sub-panel as well as new ideas on impact compared to the 2008 RAE. This would give more time for development and testing of templates for submission of information and templates for sub-panels to use as well as to possibly to perform some pilot studies on impact assessment. Moving the assessment to a 6 year cycle would also save overall costs.

There needs to be more consultation with sub-panels to provide information in the form they require to minimise their work-load. This did not happen for the 2008 RAE. In addition more time is needed for the assessment process which was very rushed in 2008.

Expert review

Question 1: Do you agree that selection of staff by institutions is the most workable approach, and that there should be transparency about the proportion of staff selected?

We agree. Any other approach is likely to lead to game playing, e.g., via re-badging, and raise complex HR issues for institutions and individuals, for example around staff whose primary duties are not research.

Question 2: How should we clarify the definition of eligible staff?

The eligibility criteria used in RAE 2008 were unsatisfactory in a number of respects: they were ambiguous; did not adequately capture the presumably intended concept of *independent researcher*, especially in the case of research staff; and they encouraged game playing. We urge HEFCE to consult widely on this issue and to take legal advice in order to ensure that the criteria are legally robust, unambiguous, accurately describe “independent researcher”, take into account early career staff, respect equality and diversity considerations, and obtain widespread consent. Some alignment with HESA categories might be helpful to institutions. We have a diverse range of suggestions that we would feed into such a consultation.

Question 3: How would expert panels make use of information about the proportion of staff submitted (for example, as part of the ‘environment’ element)? See above

The proportion of staff submitted should be one of the assessment criteria under environment. In particular, the ratio of submitted to eligible staff, in conjunction with other relevant information, would inform judgements about the research culture of the department and its sustainability. It would

also qualify statistics about the number of postgraduate students and the research income per FTE of eligible staff. The main objective is to allow panel members to determine 'game playing' whereby Departments submit a small percentage of staff who are presented as co-supervisors of all research students and co-investigators of all research grants. It also assists in determining whether there is the critical mass required for a sustainable research environment.

Question 4: On balance, do you agree that outputs should be attributed to the author rather than credited to the HEI at the time of publication?

Yes, it makes the process easier to manage, and it encourages mobility and competition. Also the QR funding should go to where the people are, not where they were.

Question 5: Do you agree that research by Category C and D staff should not be included in the REF?

In UoA 23, the numbers of Category C and D staff submitted to RAE 2008 were very low and had no significant impact on the final profiles. The institutional cost in preparing these cases is not worth paying, given this low impact. Of course, in other UoAs, such as medicine, the impact *is* significant and justifies inclusion.

Question 6: How else could Category C and D staff inform the assessment in disciplines where their role is important?

They could be incorporated into both the environment and impact stories, since these are often people funded by external agencies that will be promoting collaboration and/or commercialisation.

Question 7: Should panels be able to vary the number of outputs per researcher to be submitted, depending on the discipline? If so, what should be the minimum and maximum number?

Four outputs per person worked well for UoA23 in RAE 2008, representing a good balance between accuracy and workload. The exceptions for early career researchers also worked well. We see no reason to depart from these arrangements for the REF.

Submitting a smaller number of outputs could lead to profile grade inflation compared to RAE2008, assuming people accurately select their best outputs for submission.

Question 8: Would double-weighting certain output types be a useful way to reduce the burden of assessment (particularly in disciplines where citation indicators are limited or not available)?

Books and similar major outputs did not constitute a large proportion of the outputs submitted to UoA 23 in RAE 2008. We agree that such outputs could be double-weighted. The criteria for double weighting need to be clearly formulated, so that institutions submitted them do not run the risk of not meeting their output quota because outputs they expected to be double weighted were single weighted.

Question 9: Are there other ways of reducing the need to assess all outputs in detail (particularly in disciplines where citation indicators are limited)?

There is no consensus of other ways of reducing workload, so the number of sub-panel members should be proportional to the number of FTEs submitted to the sub-panel.

Question 10: Should the descriptors for 1 to 4* be reviewed? Should panels have discretion to vary them?*

There is a need for some revision of the criteria as 2*, 3* and 4* all related to international quality research. Reference to international quality should be removed from 2*. 2* quality could, for instance, require national quality in originality, rigour *and* significance, whereas 1* quality might require national quality in only one of these criteria.

Question 11: How can international benchmarking and the role of international members be enhanced (particularly in disciplines where citation indicators are limited)?

International members did not make a useful contribution in RAE 2008. If they were allocated to sub-panels, they might play a significant role in quality checking the provisional assessments, e.g., by sampling at borderlines.

Question 12: Do you agree that statements about user significance should be provided for outputs with particular significance to users?

UoA 23 took user significance into account in assessing outputs in RAE 2008. This worked well and this practice should continue in REF. Other sub-panels should decide whether it is appropriate for them.

Environment

Question 1: Which are the key issues that the environment element should focus on?

UoA 23 produced a set of criteria for assessing environment in RAE 2008. These are included as appendix A. They worked very well, making the assessment process efficient, uniform and fair. We think they are generic and commend them to other sub-panels. If they are adopted, then institutions should be required to organise their environment submissions under these headings, in order to simplify the submission and assessment process and ensure that nothing is overlooked.

Question 2. Should the wider environment or whole department be assessed, rather than focusing on selected staff?

Most of the criteria in Appendix A refer to department-wide attributes. Using these criteria *entails* assessing the whole department. Some of the statistics in RAE 2008 were presented on a per Category A basis, but this led to game playing, e.g., with research students and research assistants being reassigned from unsubmitted to submitted staff, so that some submitted staff had unrealistic allocations. Presenting these statistics at a whole department level would prevent such game playing and be easier for institutions to compile.

Question 3. Would data about the profile of (all eligible) staff be valuable to the assessment?

Yes, for the reasons given in the answer to question 2. In using their judgement to assess the significance of these statistics, sub-panels can take into account the ratio of submitted to eligible staff and the textual commentary describing the research structure of the department.

Question 4. Would data about PGRs and research income that is readily available from HESA be sufficient for the purposes of the REF?

Yes, for the reasons given in the answers to questions 2 and 3 above.

Question 5. What other key indicators should be included (as part of a common family of indicators

for research environment)?

See Appendix A

Question 6. What are the key questions that we could ask about research environment to be included in a pro forma?

See answer to question 1 above. Submissions should be organised under the headings given in appendix A.

Question 7. Should research environment be assessed using a single point scale rather than a profile?

No, a profile should be used. The criteria in Appendix A work well for forming a profile: scores under each criterion and combined, possibly with weights, to give the overall profile. A single point assessment would distort the final profile.

Question 8. What should the criteria and descriptors be for the assessment of environment?

We suggest criteria in Appendix A. The descriptors of the * rankings for outputs do not work well for either environment or impact and should be revised. We suggest simple, generic descriptors, such as: Excellent, Good, Satisfactory, Poor, Absent.

Question 9. Should we consider assessing environment at a broader (main panel) level?

No. The assessment of many aspects of environment are subject specific, so need to be assessed at sub-panel level.

Impact

Question 1: Do members agree that ideally the REF should take a broad overview of the breadth and depth of impacts achieved by a department.

Yes. Impact can take many forms, especially in such an inherently interdisciplinary subject as Computer Science and Informatics. Any attempt to prescribe the forms of impact will potentially exclude unanticipated but valuable forms. Assessment should take into account both the range of impacts and the penetration of these impacts.

Note, in particular, that foundational discoveries often take a long time to transform into societal/economic impacts, and the route can be long and convoluted. The initial impact may be on academic researchers and students both within and outside the discipline and the institution. These initial impacts must also be eligible for assessment.

Question 2: What key questions should the narrative statement address?

Key questions may be valuable in structuring the narrative to make it pertinent and easier to assess. Care must, however, be taken to ensure that these key questions are generic and do not exclude unanticipated but valuable forms of impact (see question 1). Key questions could include: the source of the original idea; the form of the impact; the evidence of impact; the target of the impact; anticipated future impact.

Question 3: What could be key indicators of impact and how would they inform the assessment?

Examples of impact indicators are given in Appendix B.

Question 4: How much case study evidence would be required to inform the assessment?

We anticipate a case study per main area of research activity. An overall size limit must be imposed on the narrative, that is proportional to the size of the department.

Question 5: How should the challenges of time lags and attribution be addressed?

Credit should be given only for impact that happened during the census period and that can be attributed to submitted staff. However, the case study narrative could also describe any originating research and its previous impact before the census period, and could mention the contribution of staff who are no longer members of the department. No time limit should be imposed, provided the history is useful in placing the current impact in context.

Question 6: What should the criteria and descriptors for impact be?

Some suggested computer science and informatics impact indicators are listed in Appendix B. The descriptors for the * rankings should be the same as the generic ones we suggested for environment in question 8 above.

Question 7: Should impact be assessed using a single point scale rather than a profile?

No. Only a profile can capture the full range of indicators and case studies. For instance, a few examples of 4* impact must not be diluted by averaging these with more pedestrian impact. Exactly how these profiles are calculated is subject specific and must be determined by sub-panels.

Question 8: At what level of subject granularity should impact be assessed?

Impact is very subject specific and must be assessed at sub-panel level. In particular, computer science and informatics is inherently interdisciplinary with a very broad range of potential impacts. No main panel would be equipped to deal with this kind of diversity, but computing researchers are very familiar with doing so.

Question 9: How should research users be involved in the process?

Research users should be recruited to sub-panels, where they should play a similar role to international members, i.e., they should be used to sanity check the assessments of the sub-panel, especially on impact, but should not be expected to undertake the huge workload of output assessment. Few users could justify to their employers taking on such a workload, so insisting on this would make them very difficult to recruit.

The three elements of assessment

Question 1: Is the proposed coverage of impact workable across all disciplines?

We support the plans for coverage of impact, except those for academic impact. Restricting to academic impact on *other* disciplines is not good enough, especially for an inherently interdisciplinary subject, such as CS&I. Impact chains in CS&I are often long, complex and hard to track. Restrictions on which parts of these chains count as impact, run the risk that some important impacts will be overlooked. Impact chains link both to and from other disciplines, but also link different parts of our discipline, e.g., theoreticians and system builders. Such links are analogous to

mathematicians influencing engineers. It's not clear why one should count but not the other. We have provided a detailed proposal on what can be assessed as impact in Appendix B

Question 2: Where is further clarity needed in terms of the relationship between the three elements?

There are potential slight overlaps between impact and the 100 word significance statement relating to each output. However, the significance statement relates to specific outputs and impact is likely to relate to a set of outputs over a period of time which may be earlier than the assessment period. Any overlaps can be dealt with by sub-panel members' judgement.

Question 3: Are standard weightings between the three elements feasible or is some flexibility required between UoAs or main panels (and if so, why)?

CS&I would support a standard weighting in order to get consistency between sub-panels. We think that a 60:20:20 ratio gives appropriate weight to the different elements of the assessment. We see no advantage in rounding in producing final profiles, and suggest that only unrounded profiles be used.

Question 4: Should impact and environment produce sub-profiles or single point ratings, and how should they be combined?

Profiles are necessary in order to balance appropriately the different elements of impact and environment and to allow pockets of excellence to be highlighted and not be lost in the averaging. Our proposals on criteria for impact and environment assume sub-profiles.

REF panels

Question 1: Do members agree there should there be fewer UoAs and less variation in the volume of work covered by each UoA? How can this be achieved?

To ensure an equitable and practicable assessment workload, there needs to be less variation in the volume of work covered by each *sub-panel member*. Sub-panel membership should be proportional to the number of FTEs submitted to the sub-panel. This was not the case in RAE 2008. To minimise institutional burden, there needs to be good correlation between the remits of UoAs and typical departments, i.e., to minimise the number of departments submitting to several UoAs or submitting jointly with other departments. This puts a limit on the potential to reduce the number of UoAs. Otherwise, sub-panels will become unmanageably large and institutions will have the burden of many joint submissions.

Fortunately, CS&I departments have a large measure of uniformity, making CS&I a natural UoA and sub-panel, as in previous RAEs. The extension, in 2008, to include informatics was successful, simplifying issues of relevance, while remaining within the competence of the sub-panel. CS&I is, however, a very large UoA, whose sub-panel size in 2008 was too small, putting a very heavy workload on the sub-panel members. CS&I cannot sensibly be split nor combined with another UoA. The field is too closely intertwined to be separable, and already too large to be merged.

We do not think that sub-panel workload issues can be successfully addressed by outsourcing to external reviewers. Our experience of external reviews in previous RAEs suggests that they are not

well calibrated, so that their gradings are typically not reliable.

Question 2: Do members agree the areas outlined at paragraph 16 should be consistent across the exercise as a whole?

We strongly support this. In particular, we commend the environment criteria and working methods that CS&I developed and successfully used in RAE 2008, which are included as Appendix A. We think these are generic and simplify the assessment process. We also commend our proposed criteria for impact in Appendix B

Question 3: How can the REF support consistent application of internationally benchmarked standards across the panels?

No such standards exists for computing and it would be difficult to develop them.

Question 4: Are the issues outlined in paragraph 18 the main areas where main panels should achieve consistency?

Yes, we agree that this list covers the main issues.

Question 5: In which areas do differences in the nature of research between individual UoAs justify and require flexibility between individual sub-panels to vary the assessment process?

We agree with this list, but would go further and ask for bibliometric data from Google Scholar for CS&I. Google Scholar is not perfect, but compared with other sources of bibliometric data, has far better coverage of our discipline. It is in widespread and standard use in CS&I. It was widely used to inform judgement by the CS&I sub-panel in RAE 2008, and will be used by REF sub-panel members. Having this information provided will reduce the workload on hard-pressed sub-panel members and ensure uniformity of the data.

Question 6: How should the configuration of main panels be revised for the REF?

The majority of our working group think that CS&I would fit most naturally under an Engineering main panel. Although CS&I uses a diverse range of methodologies: scientific, mathematical and engineering, this very diversity and our dissemination practices fit Engineering better than other potential main panels. One of our members dissents from this view and would prefer a main panel covering a broader range of STEM disciplines, including engineering, since this would reflect all the main methodological approaches in our field.

Question 7: Are there any key issues around nominations processes or panel membership that should be revised for the REF?

The nomination process of sub-panel and main panel members has worked well for previous RAEs and should be retained.

International members should be at sub-panel level, where they can be used to sanity check the final assessments and calibration, especially of the outputs. Their role at main panel level is less clear, except to report on their sanity checking, which they can do by ad hoc main panel attendance.

Specialist advisers have not proved helpful in CS&I and can be dispensed with. In CS&I we're used to dealing with a wide variety of methodologies, and find as many within our field as outside it, so usually find it unnecessary to seek specialist advice. Where it has been used, the lack of calibration of the advisors makes their assessments of questionable value

It is difficult for research users (typically industrials) to justify to their bosses taking on a huge workload, such as is involved in output assessment. This makes it hard to recruit the best people. They should be used only in a sanity checking role, similar to the international members, but with the main emphasis on impact assessment

UKCRC REF Working Group Membership

Prof. Alan Bundy, University of Edinburgh, <a.bundy@ed.ac.uk> (Convenor).

[Prof. Nick Jennings, University of Southampton, <nrj@ecs.soton.ac.uk>](mailto:nrj@ecs.soton.ac.uk).

[Prof. Cliff Jones, Newcastle University, <cliff.jones@ncl.ac.uk>](mailto:cliff.jones@ncl.ac.uk).

[Prof. Ross D. King, Aberystwyth University, <rdk@aber.ac.uk>](mailto:rdk@aber.ac.uk).

[Prof. Ursula Martin, Queen Mary University of London, <Ursula.Martin@dcs.qmul.ac.uk>](mailto:Ursula.Martin@dcs.qmul.ac.uk).

[Prof. Tom Rodden, University of Nottingham, <tar@cs.nott.ac.uk>](mailto:tar@cs.nott.ac.uk).

[Prof. Morris Sloman, Imperial College, <m.sloman@doc.ic.ac.uk>](mailto:m.sloman@doc.ic.ac.uk).

Appendix A Environment Assessment

We anticipate that the following template could be modified or adapted to suit submission of environment information for specific subjects. Maximum word length of responses are indicated but these may be subject specific.

1. Infrastructure, facilities and administrative support for research
 - Existing infrastructure and facilities for research e.g. specific laboratories or programmer support for software tool development and evolution (200 words)
 - New institutional investment (200words)
 - Administrative support in terms of generating and managing research proposals (100 words)
2. Translating Research into Impact
 - Institution/Departmental support to enable the translation e.g. innovation and consultancy support, training, exhibition or conference organisation support, time off for attending meetings. (300words)
 - Specific training if appropriate. (100 words)
3. Arrangements for developing and supporting staff in their research
 - Support provided for early career researchers with respect to teaching load, specific finance, training etc. (300words)
 - Mentoring and management of research grant proposals e.g. internal reviewing , training (200words)
 - Research team management/supervision training (100 words)
4. Research Student Training
 - Research degrees (PhDs) awarded /FTE (Standard Analysis, RA3)
 - Current numbers of PhD Students
 - Research student training and support for conferences attendance etc. (300 words)
 - Arrangements for monitoring PhD students, encouraging completion (300 words)
5. Research income: funding strategy, amount received and sustainability
 - Total income per FTE (Standard Analysis)
 - Composition of income
 - Trajectory of income
 - Explanation of income variations over assessment period (200 words)
 - Research Assistants per FTE (Standard Analysis)
 - Studentships per FTE (Standard Analysis)
 - Description of funding sources for studentships (200words)
6. Credibility, vitality and sustainability of research organisation
Free text: 100 words per FTE submitted
 - Research Group structure and critical mass
 - Credibility, vitality and sustainability
 - Strategic vision
 - Leadership of discipline

Evaluation of Environment

Ratings 0-4* relate roughly to criteria no-evidence/none, low/adequate, good/moderate, very good/significant, excellent/substantial.

Each of the 6 main heading is allocated 10 points to be distributed between 0-4* 'boxes'. Different weightings can be allocated to each headings if appropriate e.g. 5 & 6 could be allocated double weighting.

The information from the bulleted points is used to make a subjective judgement for each main heading. An evaluation spreadsheet with tick boxes for each 0-4* ratings for every bullet point would aid the assessor in allocating points to each rating.

For example consider a UOA with mostly young staff, low numbers of PhDs completions, high numbers of current PhD students, moderate research student training program but very good system for monitoring progress and ensuring completion. Heading 4 ratings might be: 4 points in 1*, 2 points in 2*, 4 points in 3*

The spreadsheet would generate an Environment Profile from the points for each heading.

Environment for each UoA was evaluated by 3 assessors using the above method in panel F23, and we found considerable consistency between assessors.

Appendix B Impact Assessment

The following exemplar indicators may need to be adapted for each subject.

1. Industrial collaboration
100 word max descriptions of x case studies of industrial collaborations
where $x = n/10$ for n FTE staff members.
2. Interdisciplinary collaboration
100 word max descriptions of x case studies of interdisciplinary collaborations
where $x = n/10$ for n FTE staff members.
3. Impact Case studies
Examples include spin outs, tracing IPR licences into products, policy advice to government, widely used software package or tools, exhibitions, media programmes etc.
250 word max descriptions of x case studies of economic, social or academic impact where $x = n/10$ for n FTE staff members.
4. Impact & esteem indicators
Including advisory board membership; leadership roles in industry, commerce, research councils, learned societies; conference programme chairs; tutorials at conferences; membership of selective academies; journal editorships; membership of funding panels; foreign institution advisory boards; prizes
1.5 n indicators for n FTE staff members and a maximum of 3 indicators for a single individual. Must include full name of organisation and up to 20 word descriptions for those which may not be known by assessors.
Some panels criteria could specify range of different type of indicators eg. no more than y% academic esteem.
5. Overall Citation Rates
Overall citations rates relating to the department as a whole could be included e.g. taking top x cited papers for each person submitted to give a profile of citation counts for the department as indicated in the bibliometric pilot.

Evaluation of Impact

Ratings 0-4* relate roughly to criteria none, low, moderate, significant, substantial impact or esteem.

If the above impact indicators are provided to panel members in the form of a spreadsheet with each indicator, including text description as a row, it would be comparatively easy to assign a rating 0-4* for each indicator and hence generate an impact profile.

For 4 above (Impact and esteem indicators), the name of the individual is also required for auditing.

It would be possible to assign different weights to the various indicators e.g. indicators in 1, 2 and 3 may be given a higher weight than in 4.