REF Expert Advisory Groups

Summary of discussion from meetings

Introduction

1.   In December 2008, HEFCE (on behalf of the four UK Higher Education funding bodies) established Expert Advisory Groups (EAGs) to help develop proposals for the Research Excellence Framework (REF). Membership of the EAGs was drawn from RAE 2008 panels, research councils, users of research and other organisations. The terms of reference and list of members is at **Annex A**.

2.   The EAGs met three times between February and June 2009 to discuss the following issues:

* Experiences of the 2008 Research Assessment Exercise (RAE) and key features of the REF
* The use of bibliometrics
* Expert review of outputs
* Environment
* Taking account of impact in the REF
* Equalities issues
* Combining the three elements of assessment
* REF panels

3.   Each meeting consisted of a combination of plenary discussions and breakout sessions. Each breakout session comprised 10-15 members, one of whom chaired the discussions, and a note-taker from HEFCE. The breakout groups were mixed in terms of subject discipline, except for the discussions about REF panels and the three elements of assessment, where they were grouped into broad subject areas.

4.   Papers were circulated in advance of each meeting to introduce the issues and pose questions for discussion in the breakout sessions. The papers themselves have not been included in this document as they provided information about developments which have since been superseded by the consultation document itself. The questions we posed are included in the annexes of this document.

5.   In advance of the discussion on bibliometrics, members were given an overview of the results of the bibliometrics pilot and access to some data relating to their specific disciplines.

6.   This document provides a high level summary of the discussions at the EAGs on each of the areas listed above. The questions posed on each issue and a full summary of the discussions in the breakout groups can be found at **Annexes B to H**.

Key points

<u>Experiences of the 2008 RAE and key features of the REF</u>

Scope and criteria

7.   There was felt to be some intention around higher education institutions' (HEI) selection and recruitment of staff for the 2008 RAE to improve quality profiles (not necessarily to maximise funding). Many members felt the REF should be more inclusive and avoid HEI selection if possible, but also identified a number of problems with a 'universal' approach.

8.   There were mixed views on whether outputs should be credited to the author (as at present) or whether the employing institutions should get the credit.

9.   Members queried whether it is worth the effort of including Category C staff (except in some specific disciplines such as Medicine where particular considerations apply).

Panel structure and configuration

10.   RAE main panels were felt to be effective in ensuring consistency within their main panel. There was no robust process intended to ensure consistency between main panels.

11.   The RAE sub-panels were felt to cover coherent fields of research, but there was a wide disparity of workloads between panels, and the workload for some was very heavy, particularly where a large number of outputs had to be read.

12.   Given the demands of reviewing outputs and the need for experts to interpret bibliometrics, members felt there is limited scope for moving to fewer, broader panels. Some members felt, however, that environment or other elements (not outputs) could potentially be assessed at a broader level (similar to RAE main panels), and that this could make the effort of assessing these elements more proportionate.

Quality profiles

13.   Members agreed that the quality profile used in the 2008 RAE 2008 was an improvement on the single point ratings system used in 2001: quality profiles were more discriminating and revealed 'pockets of excellence'. Members agreed that profiles should be retained in the REF (although some queried whether outputs and other elements would need to be combined into an overall quality profile).

Outputs

14.	There was agreement that outputs were central to assessing quality in the RAE and should continue to be so for the REF.

15.	Given the need to profile outputs, members generally felt there was limited scope to sample outputs in the RAE. Some suggested that the REF should require three outputs per person rather than four, especially if the assessment period is shorter. (Note that the use of bibliometrics in the REF was not discussed at these meetings.)

Environment and esteem

16.	There was consensus that consideration of environment was useful in the RAE although there were some challenges of interpretation; and that environment should be retained in the REF with some modifications. The focus should be on sustainability and forward looking aspects. The narrative evidence should be more structured than the present RA5. This, along with relevant indicators, should inform expert judgement about environment and a mechanistic approach should be avoided.

17.	There was a widespread view that the esteem sub-profile was less useful and a distinct element for esteem is not needed in the REF (although some of the indicators of focusing on 'public good' elements and research income should be retained under a different heading).

18.	Some found it difficult to make profiles out of environment and esteem as profiles require fine-grained judgements. Some felt the criteria of 'originality, significance and rigour' and the descriptors of 1* to 4* were designed primarily for assessing outputs and were less applicable to environment and esteem. The REF should consider tailoring the criteria, ratings scale and descriptors for environment and other elements. Some suggested that while outputs should continue to be profiled, other elements could be assessed on a single point scale.

Applied research

19.	Members felt that applied and practice-based research was treated fairly where it was submitted to the RAE, but (with some exceptions) HEIs had been reluctant to submit applied research. Also, there were some difficulties assessing it against the same criteria as basic research.

Impact and engagement

20.	Members agreed or accepted that it will be important to take more explicit account of the impact of research in the REF, and that there should be a distinct element focused on impact (or impact and engagement). There should be a wide definition of impact, including social, cultural, public policy and quality of life benefits, as well as economic impact.

21.   Members highlighted several challenges to assessing impact, including time lags and problems with attributing impacts to research outputs. They discussed two broad approaches:

a.   Looking at impact that has occurred during the assessment period arising from research by the HEI over a much longer period.

b.   Looking at impact 'enablers' such as engagement with users and knowledge transfer activity.

22.   Members agreed that there should be more user representatives on the REF expert panels, with a key role in informing the assessment of impact.

Interdisciplinary research

23.   In general, members felt that panels were able to assess interdisciplinary research without difficulty, given the breadth of panel membership and use of specialist advisors. However, many noted practical difficulties with the cross-referral process.

24.   A full summary of the discussions about lessons learned from the RAE and key features of the REF is at **Annex B.**

The use of bibliometrics in the REF

25.   There was a strong consensus that bibliometrics are not sufficiently mature to be used formulaically or to replace expert review, but there is considerable scope for citation indicators to inform expert review in the REF.

26.   There was widespread agreement that the most appropriate approach is to focus on citation indicators for selected papers by the staff in each submission, rather than attempt to capture all papers.

27.   There are a number ways in which bibliometrics can be used to inform expert review to enhance the reliability of the process and in some cases reduce panel workloads; the particular ways in which the data are used could vary between panels.

28.   A full summary of the discussions about bibliometrics is at **Annex C**.

Expert review of outputs

29.   There was widespread agreement that selection of staff by institutions is the most workable approach. Members made various suggestions about how the definition of eligible staff could be clarified. Most agreed that there should be more transparency about the proportion of staff selected, and that this could inform panels' assessment of a

department's research environment, although concerns were raised about how this could work in practice.

30.   There was widespread agreement that on balance, outputs should be attributed to the author rather than credited to the HEI at the time of publication.

31.   Most agreed that research by Category C staff should not be included in the REF, except for subjects where their contribution is of particular importance (such as medical subjects).

32.   Several options were explored around the number of outputs that should be submitted per researcher, and how the burden on panels of reviewing large volumes of outputs could be reduced. There was no clear consensus about a preferred approach, although many felt that four outputs per person was 'about right' and around half the groups felt that double-weighting certain types of outputs could be a workable way of reducing burden for panels.

33.   Many agreed that statements about the 'user significance' of outputs should be included where relevant, but there were concerns about how this would overlap with the impact element of REF.

34.   On the whole, members felt the descriptors for 1* to 4* should be reviewed, in particular to distinguish more clearly between international levels of excellence.

35.   A full summary of the discussions about expert review of outputs is at **Annex D.**

Environment

36.   Members felt the environment element should focus on sustainability, staff development and strategy, and identified a range of indicators that could be used to inform this. Some felt the assessment should consider the whole 'department' rather than just selected staff; and in principle agreed that readily available data on research students and income from HESA should be used.

37.   The majority felt that environment should continue to be assessed using a profile rather than a single point scale, and that it should be assessed at sub-panel rather than main panel level.

38.   A full summary of the discussions about environment is at **Annex E.**

Taking account of impact in the REF

39.   Members generally agreed that the REF should take a broad overview of the breadth and depth of impacts achieved by a department. This should include benefits to

the economy, society, public policy, culture, health, wellbeing and quality of life. There was a widespread view that academic impact should also be included, particularly where the impact extends beyond the department's field of research to influence other disciplines. Some members also felt strongly that 'public engagement' should be included.

40. There was support for the suggested approach of using a narrative statement supported by case studies and indicators, although some expressed concern about the additional workload involved.

41. In general members agreed that the REF should look at impacts that have become evident during the assessment period arising from research undertaken over a longer time frame; further work would be needed around time limits for this, how attribution would work in practice, and how claims could be substantiated.

42. Overall, members preferred the option of assessing impact using a profile rather than a single point scale. There was also general agreement that impact should be assessed at sub-panel rather than main panel level.

43. Many felt that research users should have a key role in validating case-studies and assessing the impact element, but that they should not be expected to read large volumes of outputs.

44. A full summary of the discussions about impact is at **Annex F.**

Equalities issues

45. In general, members agreed the guidance and training for panels and the guidance given to HEIs was sufficient in RAE 2008, but that more could be done to ensure greater consistency and clarity in the REF. In particular there should be more consistent treatment of early career researchers (ECRs) across the exercise.

46. There was support for the idea of a central team to develop guidance and possibly adjudicate claims of special circumstances, rather than these being assessed individually at sub panel level.

47. A full summary of the discussions about equalities issues is at **Annex G.**

Combining the three elements of assessment

48. The groups had mixed views about appropriate weightings for each element of assessment. Many supported standard weightings across all panels, though different suggestions were made about what these should be. The range of suggestions was as follows:

- Outputs: 60%-75%

- Environment: 10%-20%
- Impact 10%-20%

49.   Most felt that both the impact and environment elements should be scored using a profile rather than a single point scale, although there should be a consistent approach across the exercise in terms of how they are constructed. Many also emphasised the importance of having distinct level descriptors for environment and impact.

50.   A full summary of the discussions about the three elements of assessment is at **Annex H**.

REF panels

51.   On the whole, members agreed that there should be fewer units of assessment (UOAs) in the REF, fewer fluid boundaries between UOAs and less variation in the volume of work covered by each panel.

52.   Members agreed that the REF should achieve greater consistency across all UOAs than in the RAE and that variation between individual UOAs should only occur where they are justified by the nature of the discipline.

53.   The great majority agreed that many aspects of the overall assessment framework should be standardised across all UOAs, including the criteria and definition of quality levels, the range of weightings, the definition of eligibility, treatment of special circumstances, maximum number of outputs per person, the role of panels and the method, format and core data of submissions.

54.   There was strong support for the proposal of fewer, broader main panels than in the RAE 2008. On the whole, there was also broad agreement that, assuming main panels can be configured to cover related or cognate groups of UOAs, the REF should aim to achieve consistency of criteria, working methods and assessment standards within each main panel. Members made suggestions about how panels could operate to help support this.

55.   There was some support for sharing panels' workload with reviewers who are not full members of the panel, although some concerns were raised about how this would work in practice.

56.   A full summary of the discussion about REF panels is at **Annex I**.

Annex A

EAG terms of reference and membership

Introduction

1.   On behalf of the four UK Higher Education (HE) funding bodies, HEFCE has established Expert Advisory Groups (EAGs) to help develop proposals for the Research Excellence Framework (REF).

2.   Membership of the EAGs has been drawn from RAE 2008 panels, Research Councils, users of research and other organisations.

Purpose of the groups

3.   The EAGs will meet during the first half of calendar year 2009, to help inform the development of proposals for the REF. HEFCE and the UK HE funding bodies will then consult the sector on the proposals during the latter half of 2009.

4.   The EAGs will provide advice on the overall design of the REF, as a unified framework for assessing the quality of research across all disciplines and all types of research. This will include advice on:

- the approach to expert review of outputs
- the role and use of bibliometric indicators of research quality
- the choice and use of other quantitative and qualitative indicators within the framework
- how to take account of social, economic and public policy impacts within the REF
- the configuration of expert panels and their roles in the assessment process
- how to accommodate key differences between subject groups within a unified framework for assessment
- other aspects of the framework, to ensure that our proposals meet the objectives of the REF and can carry the confidence of the sector.

5.   The EAGs will discuss these issues at three rounds of meetings between February and June 2009. Members will be encouraged to draw on their own expertise and experience as well as consult informally with their subject communities.

6.   The EAGs will convene for a fixed term, specifically to advise on the development of REF in advance of the consultation. After the consultation exercise in 2009 has concluded, the UK HE funding bodies will establish REF expert panels to develop subject-specific criteria for assessment and conduct the REF assessment process.

7.   In addition to seeking advice from the EAGs as we develop proposals for the REF, the UK HE funding bodies will hold discussions with a range of other key stakeholders including institutional managers, other funders and users of research.

Membership

8. RAE 2008 Main Panel Chairs, the Chief Executives of the Research Councils and several other organisations have helped us to identify members with expertise in research assessment, from across the full range of disciplines and types of research activity in the UK. The membership is listed at Annex A.

Role of members

9. It is intended that members of the EAGs will provide expert input based on their expertise and experience, and by consulting informally with and drawing on the views of their wider communities. Members will be encouraged to do this through the course of their normal business, and it may also be appropriate to attend or arrange meetings with specific groups (such as learned societies) to communicate current thinking and gather views on REF development.

10. HEFCE will provide members of the EAGs with materials to help support their wider discussions, and may also publish these on our website. We envisage that in some cases HEFCE staff will work with members to consult particular groups, for example by helping to organise meetings or accompanying members to certain meetings.

Confidentiality issues

11. As described above, we intend for members of EAGs to have an open dialogue about the REF with their wider communities in order to fully engage them in the development of REF.

12. However, members will be required to keep certain information confidential and may be asked to sign a confidentiality agreement. This will be particularly important with regards to bibliometrics pilot data and other sensitive data which may be made available to members.

Focus and format of meetings

13. The initial round of meetings (in February 09) will cover key features of the REF, drawing in particular on the experience of the 2008 RAE. This round of meetings will be mixed in terms of the range of disciplines and types of research.

14. We expect two further rounds of meetings, as follows:

a. A second round of meetings in April/May 2009, to focus on the main elements of assessment. This will include detailed discussion of expert review of outputs, approaches to assessing impact, the use of bibliometrics, and the choice and use of other indicators and information.

b.   A final round of meetings in June 2009, to further define the elements of assessment, advise on how they should be combined to form overall quality profiles, consider the role and constitution of expert panels, the subject configuration, and the scope for variation between subjects.

Membership list

15.   The following is a complete list of members of the EAG.

First name	Surname	Organisation	Role in RAE 2008 or other relevant information
Margaret	Atack	University of Leeds	Chair sub-panel 52, French
Graeme	Barker	University of Cambridge	Chair sub-panel 33, Archaeology
Jonathan	Bate	University of Warwick	Member sub-panel 57, English (Nominated by AHRC)
John	Bell	University of Cambridge	Member sub-panel 38, Law
Dianne	Berry	University of Reading	Chair sub-panel 44, Psychology
Alexander	Bird	University of Bristol	Member sub-panel 60, Philosophy
Jane	Broadbent	Roehampton University	Member sub-panel 36, Business and Management Studies and 35, Accounting and Finance
Kenneth	Brown	University of Glasgow	Chair sub-panel 20, Pure Mathematics
Margaret	Brown	King's College London	Chair sub-panel 45, Education
Bruce	Brown	University of Brighton	Chair main panel O
Paul	Cammack	Manchester Metropolitan University	Chair sub-panel 47, American Studies and Anglophone Area Studies
Heather	Campbell	University of Sheffield	Deputy chair sub-panel 31, Town and Country Planning
Michael	Cates	University of Edinburgh	Member sub-panel 19, Physics
Celia	Caulcott	BBSRC	Director of Innovation & Skills, BBSRC
Karen	Cox	University of Nottingham	Member sub-panel 11, Nursing & Midwifery
Catherine	Davies	University of Nottingham	Chair sub-panel 55, Iberian
Maria	Delgado	Queen Mary, University of London	Member sub-panel 65, Drama and Performing Arts
Hastings	Donnan	Queen's University Belfast	Chair sub-panel 42, Anthropology
Ann	Dowling	University of Cambridge	Chair main panel G

John	Feather	Loughborough University	Chair sub-panel 37, Library and Information Management
Janet	Finch	Keele University	Co-chair of the Council for Science and Technology
David	Finnegan	University of Edinburgh	Chair sub-panel 14, Biological Sciences
David	Firth	University of Warwick	Member sub-panel 22, Statistics and Operational Research
Ray	Fitzpatrick	University of Oxford	Chair sub-panel 7, Health Services Research
Chris	Francis	IBM United Kingdom Ltd	Government Programmes, IBM (Nominated by CBI)
Peter	Golding	Loughborough University	Chair sub-panel 66, Culture, Communication and Media Studies
Peter	Goodhew	University of Liverpool	Chair sub-panel 29, Metallurgy and Materials
Neil	Halliwell	Loughborough University	Chair sub-panel 28, Mechanical, Aeronautical and Manufacturing Engineering
Russell	Hamilton	Department of Health	Member main panel A
Philip	Hannaford	University of Aberdeen	Chair sub-panel 8, Primary Care
Barbara	Harriss-White	University of Oxford	Chair sub-panel 43, Development Studies
Graeme	Henderson	University of Bristol	Member sub-panel 15, Pre-clinical and Human Biological Sciences
Sandy	Heslop	University of East Anglia	Chair sub-panel 64, History of Art, Architecture and Design
Carole	Hillenbrand	University of Edinburgh	Chair sub-panel 48, Middle Eastern and African Studies
Alison	Hodge	QinetiQ	University Partnership Director, QinetiQ (Nominated by CBI)
Tracey	Howe	Glasgow Caledonian University	Member sub-panel 12, Allied Health Professions and Studies
Pat	Hudson	Cardiff University	Member sub-panel 62, History
Martin	Humphries	University of Manchester	Member sub-panel 14, Biological Sciences
Jacqueline	Hunter	GlaxoSmithKline	Senior Vice President, Science Environment Development, GSK
Vivienne	Hurley	British Academy	Head of Policy Development, British Academy
Hazel	Jeffery	NERC	Head of Strategic Management, NERC
Iain	Jones	ESRC	Director of Evaluation and Impact, ESRC
Richard	Jones	University of Sheffield	Nominated by the Royal Society
Roger	Kain	Exeter University	Treasurer and Vice-President of the British Academy

Sandra	Kemp	London College of Communication	Member sub-panel 63, Art and Design
Ross	King	University of Aberystwyth	Member sub-panel 23, Computer Science and Informatics
Victor Terence	King	University of Leeds	Chair sub-panel 49, Asian Studies
Mike	Lant	Syngenta	Licensing Manager, Syngenta (Nominated by CBI)
Cristina	Lazzeroni	University of Birmingham	Nominated by the Royal Society
Judith	Lieu	University of Cambridge	Member sub-panel 61, Theology, Divinity and Religious Studies
Peter	Littlejohns	NICE	Member sub-panel 7, Health Services Research
John	Local	University of York	Chair sub-panel 58, Linguistics
David	Lynn	The Wellcome Trust	Head of Strategic Planning & Policy, The Wellcome Trust
Malcolm	MacCallum	Queen Mary, University of London	Member sub-panel 21, Applied mathematics
Antony	Manstead	Cardiff University	Member sub-panel 44, Psychology
Mary	Matthews	Blitz Games Studios	Strategy and Business Development Director, Blitz Games Studios
Peter	McGuffin	King's College London	Chair sub-panel 9, Psychiatry, Neuroscience and Clinical Psychology
Quintin	McKellar	Royal Veterinary College	Member sub-panel 16, Agriculture, Veterinary and Food Science
Andrew	McMichael	University of Oxford	Chair sub-panel 3, Infection and Immunology
Simon	McVeigh	Goldsmiths College, University of London	Member sub-panel 67, Music
Ed	Metcalfe	South East England Development Agency	Head of Science, Technology, Enterprise & Management, SEEDA
Judie	Newman	University of Nottingham	Chair main panel L
Pól	Ó Dochartaigh	University of Ulster	Member sub-panel 53, German
Denise	Osborn	University of Manchester	Chair sub-panel 34, Economics and Econometrics
David	Otley	Lancaster University	Chair main panel I
Anthony	Payne	University of Sheffield	Chair sub-panel 39, Politics and International Studies
Jeremy	Pearson	King's College London	Deputy chair sub-panel 1, Cardiovascular Medicine
Alan	Penn	University College London	Chair sub-panel 30, Architecture and the Built Environment

Fabien	Petitcolas	Microsoft Research	Head of Intellectual Capital Development & Community, Microsoft Research (Nominated by CBI)
Mike	Pidd	Lancaster University	Chair sub-panel 36, Business and Management Studies
Chris	Pollock	BBSRC	Chair sub-panel 16, Agriculture, Veterinary and Food Science
Barry	Potter	University of Bath	Member sub-panel 13, Pharmacy
John	Rea	DEFRA	Strategic research partnerships and evidence assurance, DEFRA
Clare	Reddington	iShed and The Pervasive Media Studio	Director, iShed and The Pervasive Media Studio
Keith	Richards	University of Cambridge	Chair sub-panel 32, Geography and Environmental Studies
Rick	Rylance	University of Exeter	Chair sub-panel 57, English
Jeremy	Sanders	University of Cambridge	Chair sub-panel 18, Chemistry
John	Scott	University of Plymouth	Chair sub-panel 41, Sociology
Paul	Seawright	University of Ulster	Member sub-panel 63, Art and Design
Alan	Silman	Arthritis Research Campaign	Chair sub-panel 6, Epidemiology and Public Health
Julius	Sim	Keele University	Chair sub-panel 12, Allied Health Professions and Studies
Patrick	Sissons	University of Cambridge	Deputy chair sub-panel 4, Other Hospital Based Clinical Subjects
Morris	Sloman	Imperial College London	Member sub-panel 23, Computer Science and Informatics
Sue	Smart	EPSRC	Head of Performance & Evaluation, EPSRC
Michael	Smith	Loughborough University	Chair sub-panel 50, European Studies
Phil	Sooben	ESRC	Observer, main panel J
Steve	Sparks	University of Bristol	Chair sub-panel 17, Earth Systems and Environmental Sciences
Pauline	Stafford	Emeritus Professorship at University of Liverpool	Chair sub-panel 62, History
John	Stageman	AstraZeneca	Member main panel C
Andrew	Stark	University of Manchester	Chair sub-panel 35, Accounting and Finance
James	Stirling	University of Cambridge	Member sub-panel 19, Physics
Robert	Stout	Emeritus Professorship at Queen's University Belfast	Chair main panel B

Peter	Taylor-Gooby	University of Kent	Chair sub-panel 40, Social Work and Social Policy & Administration
Geof	Tomlinson	University of Sheffield	Chair sub-panel 25, General Engineering and Mineral & Mining Engineering
Liz	Towns-Andrews	STFC	Director of Knowledge Exchange, STFC
Nigel	Vincent	University of Manchester	Chair main panel M
Robin	Wallace	University of Edinburgh	Nominated by EPSRC
Jeremy	Watson	ARUP	Global Research Director, ARUP
Simon	Watts	Thales UK, Aerospace Division	Vice-President, Deputy Scientific Director, Thales UK (Nominated by CBI)
Shearer	West	AHRC	Member sub-panel 64, History of Art, Architecture and Design
Michael	Whitby	University of Warwick	Chair sub-panel 59, Classics, Ancient History, Byzantine and Modern Greek Studies
Colin	Whitehouse	STFC	Deputy Chief Executive, STFC
Michelle	Wickenden	NERC	Team leader for business performance management, NERC
Paul	Wiles	Home Office	Chief Scientific Adviser, Home Office
David	Williams	University of Southampton	Chair sub-panel 10, Dentistry
Clyde	Williams	Loughborough University	Chair sub-panel 46, Sports-Related Studies
Steve	Williamson	University of Surrey	Chair sub-panel 24, Electrical and Electronic Engineering
Sharon	Witherspoon	The Nuffield Foundation	Member main panel J
Andrew	Wyllie	University of Cambridge	Chair sub-panel 5, Other Laboratory Based Clinical Subjects
Lawrence	Young	University of Birmingham	Deputy chair sub-panel 2, Cancer Studies

Annex B

Reflecting on the 2008 RAE and the key features of the REF

Reflecting on the 2008 RAE - Questions

1.   EAG members were asked to discuss the following questions about the 2008 RAE:

a.   **Eligibility criteria and selection of staff**. What implications did the eligibility criteria and definitions of staff categories have for the assessment and its outcomes? What implications did institutional approaches to staff selection have for the conduct of the assessment and its outcomes?

b.   **The configuration of Units of Assessment and the two-tier panel structure**. To what extent did the Units of Assessment cover coherent fields of research and provide for manageable workloads for panels? How effective was the two-tier structure in ensuring consistency across the exercise? Were there any unintended consequences?

c.   **The use of a quality profile**. What implications did this have for the conduct of the assessment? How effective was it in meeting the desired aims of eliminating the averaging effect of the previous RAE grades, and highlighting pockets of excellence? Did it encourage greater inclusiveness of selection by HEIs?

d.   **The introduction of sub-profiles of outputs, environment and esteem**. What was the range of approaches used for reviewing outputs and determining the output sub-profiles? How useful were the esteem and environment indicators in informing the overall quality profiles? What challenges were there in interpreting quantitative indicators? What challenges were there in assessing narrative evidence (the RA5)? Which particular types of information were most useful in informing quality judgements? Which types of information added less value? Were there any unintended consequences of using sub-profiles?

e.   **Treatment of applied and practice-based research**. To what extent were these assessed on a level playing field, and how far did the efforts to encourage such an approach impact on the range of work submitted? How far did the 2008 RAE capture the contribution of research to public policy making?

f.   **Treatment of interdisciplinary and multi-disciplinary research**. To what extent were there challenges in assessing interdisciplinary and multi-disciplinary research? How effective were the mechanisms to cross-refer or seek specialist advice? To what extent did the guidance on these issues impact on the type of work submitted?

g.   **Promoting equality and diversity**. How were special circumstances taken into account? What were the challenges involved?

h.   **Efficiency**. How manageable were panels' workloads? What processes could have been simplified or streamlined? Was the overall timescale for the exercise workable and efficient?

Reflecting on the 2008 RAE - discussion

2.   The following is a full summary of the discussions of the 2008 RAE from across all breakout groups.

Eligibility criteria and selection of staff

3.   There were some general concerns around the selection of staff for submission. Panel members detected various practices used to improve quality profiles by submitting an elite selection of staff, recruiting staff close to the census date or appointing part-time staff who were not fully integrated into the department. There was concern that this does not provide a full or fair picture of a department's research.

4.   Many panels had found it difficult to verify whether the research of Category C staff included in a submission was focussed in the submitting department, and found the audit process complicated and time consuming. For some of these, the inclusion of Category C staff made little difference to the overall outcome, so it was felt to be a somewhat wasted effort. An alternative was suggested to reference these individuals as part of the 'environment' element. However, it was widely felt that removing Category C staff would be a serious omission for some panels, including medical panels. Some suggested that each panel could make their own decisions about whether Category C staff could be included when setting criteria.

5.   There were very mixed views on whether all research staff should be assessed or HEIs should be able to select staff in future. This is considered in Annex B.

The configuration of Units of Assessment and the two-tier panel structure

6.   Most groups felt that main panels were effective in ensuring consistent standards within their main panel areas, although some felt there needed to be a clearer definition and understanding of how the main and sub-panels should interact.

7.   Consistency within main panels was achieved more easily where the subjects were cognate, such as engineering. It was also helpful if the main panel chair worked closely with sub-panel chairs and attended sub-panel meetings. If there was variation of criteria and working methods between different subjects within the same panel, this hindered consistency across the panel.

8.   Many stated that they perceived limited consistency between main panels. Some did not have time within the timescales to feedback fully across the main panels, or in some cases felt discouraged from talking to chairs from other panels. This caused difficulties especially for cognate subjects from different main panels or for subjects which were split across main panels. Some felt an early steer about marking expectations would have helped to improve consistency across main panels.

9.   Some examples were given of where HEIs had flexibility to submit work to a choice of two or three different panels within different main panel areas. It was felt they made strategic decisions on where to submit, based on different panels' criteria or weightings. Greater consistency of approach across main panels would minimise this behaviour.

10.   Where main panels contained particularly coherent groups of sub-panels this was felt to be a strength. Some felt smaller sub-panels were overwhelmed by the bigger sub-panels within their main panel.

11.   It was felt that workload varied significantly between sub-panels. Workloads were particularly large where panels covered multiple fields and some of the larger UOAs had huge volumes of material to read, some of which was outside their immediate specialisms.

12.   Working methods also varied between panels, often depending on the scale of submissions and the types of outputs.

13.   Members reported mixed experiences of international panel members. Some panels found it harder to identify a specific role for their international members. Some found international panel members were more useful when involved at sub-panel level. Several panels found they did not have a balanced mix of expertise, given that the panels were set up before it was known what range of work HEIs would submit.

The use of a quality profile

14.   There was widespread agreement that the introduction of the quality profile had been a positive development. Many felt that it had taken away the 'cliff-edge' effect of using a single point scale, succeeded in showing the spread of quality and identifying 'pockets of excellence'. Although almost all agreed about the benefits overall, some highlighted issues associated with using quality profiles:

    a.   Many felt strongly that the cumulative rounding methodology (rounding to the closest 5 per cent) was a blunt measure that could have unintended effects. The quality profile improved discrimination, but this was lessened by the rounding methodology.

    b.   Some felt it was difficult to interpret and use profiles when comparing departments within the HEI (e.g. a profile with 20% at 4* could mean different things depending on the UOA).

c.   Quality profiles were more difficult to explain to the outside world than quality ratings used in the 2001 RAE.

d.   A few mentioned that quality profiles did not encourage inclusiveness as HEIs tried to exclude 1* or 2* work.

15.   However, now that the quality profile has been accepted, some urged HEFCE against changing the scales again. Repeatedly altering the scoring system makes it difficult to make comparisons for research management purposes.

<u>The introduction of sub-profiles of outputs, environment and esteem</u>

16.   Members discussed their experiences of using the three sub-profiles, as outlined below. Generally it was felt there should be more consistency across panels in how the three elements are defined and what evidence is used to inform them. Some believed weightings for each element varied too much, and should have been more constrained across the board. 5 per cent was felt by some to be too low a weighting for any of the elements.

Outputs

17.   It was widely agreed that outputs were the most important element and should remain so in the REF.

18.   Members discussed approaches to sampling outputs. Some had used sampling to some extent, but most felt that there was limited scope to do so. Given the emphasis on profiling the outputs, some said their panels had felt it necessary to review all of the outputs. This was different from previous RAEs when members felt they could sample the work of individuals.

19.   Members reported a range of approaches in terms of how many people read each output and the extent to which they reviewed outputs in detail. There was some agreement that if work appeared in a peer reviewed journal it was not always necessary to read the full text in detail. Others felt that there was no short-cut to reading material in detail, particularly monographs or in cases where outputs were felt to be borderline between star ratings.

20.   Some stated that it was difficult to assess how representative the outputs were of the body of research that has been carried out, as this varies from subject to subject. In some subjects, four outputs represented only a tiny proportion of the output from the period. Some felt we should revisit the question of whether the number of outputs should vary by subject.

21.   There was general agreement that textual commentary on outputs was useful, for example in highlighting the 'significance' of outputs or the specific contribution of the

author in multi-authored papers. Some expressed concern that a move to bibliometrics may mean these commentaries are no longer used.

Environment and esteem

22.   There was widespread agreement that environment indicators were useful in informing the overall quality profile, but that there could be improvements in how this is treated in future.

23.   The large majority of members agreed that esteem was not found to be a useful sub-profile overall and need not be retained in future. Definitions of esteem varied greatly, it was sometimes unclear how it should be assessed (especially for early career researchers) and the information included was often not verifiable. However, most agreed that some indicators of esteem (especially the 'public good' elements, and research income where this was part of esteem) should be retained as part of the assessment, perhaps under a different 'heading'.

24.   A large number reported challenges in assessing narrative evidence in the RA5. Many felt RA5s were often 'creative writing' rather than factual evidence, and were approached differently by different HEIs, which made it difficult to make judgements. Many members felt they had spent too much time on assessing RA5s, especially considering the relatively low weighting for environment or esteem in some cases.

25.   In spite of this, RA5 was felt to provide useful contextual information about the department which should be retained in some form. There was strong support for the idea of using a more structured template or proforma in the future so information could be reported in a consistent way.

26.   Overall, members felt the quantitative aspect of environment was easier to interpret. Some found the data useful to sense-check the RA5 narrative. However, members did note some challenges in interpreting quantitative indicators. There was some discussion about how the income should be broken down. Also, it was felt that attempts to apply data formulaically caused some issues, and raised questions about whether all types of income should be valued equally. Peers should make judgements about what is prestigious in their disciplines.

27.   Some had found it difficult to make a profile out of environment and esteem as this requires too fine-grained judgements; and some panels took different approaches to doing this. Some suggested it would be easier to produce a rating on a single point scale or use fewer star ratings for these elements (in contrast to outputs, which were easier to profile). Some suggested the criteria and descriptors for 1* to 4* were too loose or did not apply well to environment and esteem.

28.   Some expressed concern that sub-profiles could distort the overall quality profile: a high score for environment or esteem could make a big difference overall, even if it had been given a relatively low weighting. Some suggested that environment favoured bigger

HEIs, and could cause particular problems when a high score turned out not to be backed up by outputs. Also, not knowing what proportion of staff were selected meant the panels could have a skewed view of the environment overall (e.g. if 40 students were attached to one research active staff who was the total of one submission). This is particularly problematic if the data is applied formulaically.

29.   Members of the panel felt that it was important that environment was a forward-looking element of the assessment which allowed them to take a judgement on the potential of a department, given that the outputs are essentially retrospective. The department's strategy and staffing policy (including succession planning) were key parts of environment.

Treatment of applied and practice-based research

30.   The large majority of members felt that only a relatively small quantity of applied and practice-based work had been submitted in RAE 2008 (although this was not the case for all panels). Many attributed this to concern amongst HEIs that that applied research would not be regarded as favourably as academic work, in spite of panel guidance to the contrary.

31.   A small number of panels stated that a good proportion of applied research was submitted. Some noted that where this was the case, it was often of world leading quality. Overall, it was felt that the best submissions included a portfolio of work with different types of outputs.

32.   Members agreed that where applied research was submitted, it was treated fairly. At the same time, they identified some challenges in assessing applied research against the established criteria:

    a.   Panels sometimes had difficulty simultaneously assessing the contribution to disciplinary knowledge and significance to users.

    b.   In some cases applied research could have major impact but drew heavily on previous research and therefore lacked the same degree of originality as other material.

    c.   Members also reported difficulties in assessing significance to users where there are long time scales for this to become evident.

33.   Some found that public policy research was rarely submitted as an output, but sometimes mentioned in environment or esteem. A significant number reported difficulties in assessing the impact of public policy work. This type of research tended to score less highly on rigour and originality as it often drew heavily on existing research, and some felt public policy work may have received a lower rating overall than it should. A few added that it was particularly difficult to take a view on the significance of overseas public policy work.

34.   Some reported problems with patents and suggested only those actually awarded should be eligible for inclusion.

35.   Some found the impact statement very important for describing the impact of a patent.

36.   Some gave examples of cases where practice based research was used as evidence of esteem rather than submitted as an output in its own right. On some occasions, the research that was referenced was thought to be stronger than the research submitted.

37.   Some members commented on the role of users on the panels. Overall it was felt that there were not enough user representation. Members acknowledged the difficulty of recruiting members of the user community: users would have to make a business case to justify their involvement and few could afford the time required.

Treatment of interdisciplinary and multi-disciplinary research

38.   It was generally felt that panels were able to assess interdisciplinary research without difficulties: the panels were sufficiently broad, members could assess work outside their immediate expertise, and could draw on specialist advice as required.

39.   However, a large number of members reported practical difficulties with the cross-referral process. Cross referrals were often made very late and, once returned, there was little if any time to normalise the results. It was also difficult using other panels' criteria to assess cross-referred outputs. The process was time consuming and cumbersome. It was felt that the cross referral system worked better within a main panel structure, especially where there was cross membership within a main panel group.

40.   There was widespread feeling that interdisciplinary research should be encouraged, and some concern that its treatment in the RAE failed to address perceptions about its value, and this could inhibit interdisciplinary research. Some felt HEIs had chosen not to submit interdisciplinary research, although it was often mentioned in environment statements. Some commented that very few submissions indicated where research was interdisciplinary. A few also suggested some HEIs had engaged in games-playing around where best to submit interdisciplinary work.

41.   Many felt that specialist advisors played an important role in assessing interdisciplinary research but that there were challenges in identifying and recruiting them.

Promoting equality and diversity

42.   Overall, there was strong support for the 2008 RAE approach to equalities and diversity. The new arrangements which allowed HEIs to explain special circumstances

and submit fewer outputs in some cases had significantly improved the situation for early career researchers and those with special circumstances. However, members had encountered some challenges:

> a. HEIs used different approaches to special circumstances so it was difficult to take these into account consistently. For example, different HEIs submitted two outputs or four outputs for people with the same circumstances. Some felt the criteria for the number of outputs submitted should be clearer and more consistent across panels.

> b. Some felt there were challenges in 'disentangling' information about early career researchers because of the inconsistency in approaches by HEIs. In addition, differences in the rules and criteria across panels caused difficulties for HEIs when submitting this information.

43. Some panels had different expectations of how many outputs should be submitted by ECRs, depending on the nature of the discipline. For example, in some cases it would be reasonable to expect four outputs from an ECR after working in a department for two years.

44. Some found it time-consuming to take account of special circumstances when looking at environment and esteem. The panel had to consider cases carefully to come to a fair judgement about what mitigation should be expected (e.g. expectations about travelling to conferences for those with mobility issues).

Efficiency and operational issues

45. Members gave positive feedback about the RAE secretariat, administrative support in the central team, the data collection system and the accessibility of electronic outputs on DVD.

46. Some members found the nominations process for specialist advisors was too time-consuming.

47. Members noted a significant disparity of workloads between panels. Some panels had found their workload was extremely heavy. Some felt the workload had increased from 2001 to 2008. Larger panels reported more logistical challenges.

48. A few panels felt they spent a disproportionate amount of time on very small submissions, and suggested there should be a minimum number of staff in a submission.

49. Many found the overall timescale for the exercise was very tight, with too little time available between receiving outputs and making judgements. Often there was only limited time to re-calibrate results or make adjustments to ensure consistency.

50.   Some suggested HEFCE should give more guidance to HEIs on releasing panel members from their normal jobs. Panel members had different allowances from their host institution (e.g. some had teaching relief, others didn't).

Key features of the REF - questions

51.   EAG members were asked to discuss the following questions about the key features of the REF:

52.   What are the main options for the configuration of panels and subjects in the REF?

53.   How can we change the generic criteria for research quality to account more explicitly for the impact of research (in a way that applies to all disciplines)?

54.   Do you agree that the key generic elements of assessment for the REF should be output quality, impact/engagement and environment? Are there any other key elements that should be included, bearing in mind the need to minimise burden and complexity?

55.   What are the key types of evidence and indictors that should be used to inform each of the elements above?

56.   What are the options for combining the assessment elements to produce overall quality profiles?

Key features of the REF - discussion

57.   The following is a full summary of the discussions about the key features of the REF from across all breakout groups.

Configuration of panels and subjects

58.   The members spent some time considering the scope to reduce the number of panels for the REF. Overall they agreed there was some but limited scope for fewer panels to assess outputs:

- fewer panels would require a greater number of members on each panel. This would make the operation of panels difficult, and would not reduce burden
- some were concerned about needing to combine disparate subjects and that panels would then lack coherence
- some were concerned about the dominance of larger disciplines within broad panels, and this might lead HEIs to reduce their investment in 'smaller' disciplines.

There was some debate about whether environment or other elements could be assessed independently of outputs, and if this could be done at a broader level (similar to RAE main panels). Some RAE 2008 panels had effectively assessed outputs independently of RA5s, for example by reviewing outputs first or by not allocating whole submissions to readers, and felt this approach had worked well. However, some felt that outputs should not be separated from the other elements as there was merit in seeing whether the outputs and other elements all told the same story. Some found RA5s had provided useful context when reviewing outputs in RAE 2008.

59.   Many members recognised benefits of assessing environment at a broader level. Some thought environment could be assessed at department or faculty level with information such as equipment spend, library provision etc. being key indicators for some disciplines. Depending on institutional structures, environment within an institution could potentially be assessed at several different levels (institution, faculty, department or research group). Some suggested HEFCE undertake an analysis of RAE 2008 to see how similar the environment profiles were for an institution overall and in cognate disciplines. If they are similar, this supports the case for aggregating up or assessing at a higher level.

60.   Others doubted that environment could be assessed at such a high level, and suggested discipline specific expert knowledge is needed to make an assessment of environment. Some felt that the fine-grained nature of the information and assessment was important and that aggregating up / working at a higher level would lose some of the usefulness. Some suggested that quantitative assessment in particular needs to be interpreted by experts, but perhaps qualitative assessment could be done at a higher level.

Eligibility and scope

61.   Many members felt that the REF should assess 'all staff' (rather than HEIs selecting staff) in order to provide a fairer view of the quality of research activity, a fuller picture of environment and esteem, and reduce games playing and burden on HEIs. There was some discussions about what 'all staff' would mean, but no clear consensus. Members also acknowledged various difficulties associated with this approach:
- if all 'research' staff were included on the basis of contractual status, HEIs would play games by changing peoples' contracts
- if all 'academic' staff were included, this would be disproportionate for non-research intensive institutions
- overall the exercise would reveal a higher volume of 1* and unclassified work
- even if everyone was submitted, the 'full picture' would be hampered by staff joiners and leavers within the period.

62.   There was also a range of opinion on whether papers should be credited to the HEI or the researcher.

a.   Those who thought the HEI should be credited argued that this approach would avoid 'transfer market' problems, and that it would effectively encourage HEIs to nurture new researchers.

b.   However, they did identify problems with this approach: it would be more backward-looking (whereas giving credit to current staff is more about future potential) and could lose dynamism in the system (especially for ECRs who may be looking to develop their careers by moving on).

c.   Some felt it would be technically too difficult to associate outputs with HEIs: would it be based on when the research was carried out, when the output was submitted for publication or when it was published?

d.   Most recognised that games playing could occur with either approach.

63.   Most thought that the frequency of assessment should be every 5-6 years.

64.   Some said that four outputs per submitted member of staff was about the right amount. Others suggested that fewer than four outputs could be appropriate, particularly in certain subjects, and if the assessment period is shorter (e.g. 5 years). Some thought it should be possible to have variation from panel to panel. A few commented that in RAE 2008, the fourth output tended to be weaker, and that quality would improve if only three outputs were submitted.

65.   A small number said it would be useful for contextual purposes to know what proportion four outputs represents of the total number of outputs.

Outputs, environment and impact

66.   In terms of the key elements of assessment there was widespread agreement that:
- outputs should remain the primary object of assessment
- it is important to address impact as part of the REF, but this raises a number of challenges
- environment should be retained, although with some modifications from the 2008 RAE
- esteem does not need to be a distinct element in the REF (though some of the indicators relating to 'public good' and research income should be included under a different 'heading').

67.   There was little further discussion about outputs or the use of bibliometrics, as this will be discussed at future meetings. There was much discussion about impact, engagement and environment, and some felt these could be integrated into a single element, although many accepted that politically it may be necessary to have a separate element focused on impact.

Criteria

68.   Members felt that the REF should continue to focus on quality, defined as originality, significance and rigour. It was felt that in the REF we must retain originality and rigour across the board, but could extend 'significance' to more explicitly cover both academic and wider (economic and social) impact. Members agreed that this wider impact must be broadly defined to include social, cultural, public policy and quality of life benefits, as well as economic impact.

69.   Some suggested that the criteria (originality, significance and rigour) should apply to outputs but that distinct criteria should be developed for the other elements, that focus on what panels would be looking for in terms of an excellent environment, impact or engagement. They also suggested that the ratings scale and descriptors would need to be adapted.

Impact/engagement

70.   Several general challenges were identified to assessing impact/engagement:

   a.   How to take account of the long-term impact of basic research as well as the short-term quality and impact of applied research; and the different pace at which impact occurs in different fields (e.g. longer in medicine than engineering).

   b.   The relationships between research and impact are non-linear and there are problems attributing specific impacts to specific research.

   c.   The impact of some research is not necessarily related to its quality. For example, some research that can draw heavily on existing knowledge to produce something with impact; or 'close to market' consultancy type work that can be very useful but have little intellectual contribution or international significance. Also, high-impact 'populist' research may not be quality research.

   d.   Assessing impact could lead to new types of games playing or disincentivise basic research

   e.   It was felt that it could be more difficult to find direct evidence of impact for arts and humanities disciplines than for sciences.

71.   Some groups explored approaches to assessing impact and the types of evidence that could be included:

   a.   There was some debate about the timeframe for assessing impact. Many felt that REF could look at impact that has occurred during the assessment period arising from research by the HEI over a much longer period. Some said this could be assessed at broad panel level.

b.   Some groups suggested an alternative would be to look at impact 'enablers' such as knowledge transfer activity, engagement with users and skills development. This could be evidenced through a combination of metrics (such as industrial research and consultancy income, spinouts, and so on) and narrative.

c.   One group highlighted the tensions between using coarse metrics that would miss a lot of impact and distort behaviours, and the alternative - for HEIs to produce lots of qualitative material - that would create substantial burden.

d.   Some felt it would be difficult to separate academic and wider impact, and that the 'impact' element of REF should include contributions to the discipline (such as journal editorships, creation of teaching materials, conferences and so on).

e.   Some noted that HEIs are becoming more sophisticated at measuring impact and engagement in response to research councils' requirements and HEIF. In time, internal mechanisms will be in place which could be drawn on for the REF.

f.   Some added that we should ensure that the approaches taken by the REF and the research councils must be coherent.

72.   Most agreed there should be more users on panels in the REF, although they would need to have a clear role and should not generally be expected to review large volumes of outputs. Some thought they should work alongside academics for the assessment of impact, others thought they should be more involved in setting criteria / agreeing indicators.

Environment

73.   It was generally agreed that environment should have a focus on sustainability and should be forward-looking. This should be assessed using indicators and structured qualitative evidence, including a forward looking strategy. Indicators could include:

- Numbers of PGRs and completions. Some breakdown of home/overseas PGRs may be useful. Some suggested that PGR destinations data and data on average length of time to complete PhD might be useful.
- Numbers of ECRs. In addition to numbers, there needs to be evidence of supporting career development of ECRs / integration within the department and training / development.
- Research income. A few mentioned that there should be more consistent criteria across panels about the treatment of different sources of income. Some felt that all sources of 'competitively sought' income should count equally (except perhaps RDA income). Others felt some sources of income were more prestigious than others and should be rated differently. A small number queried why overseas charity income wasn't captured in 2008 RAE.
- Investment in buildings, equipment etc.

74.   It was noted that indicators would vary between disciplines, for example research income was more illuminating for some subjects than others. Some noted the danger of indicators is that people 'work to target' and stressed the importance of using expert judgement to interpret quantitative data.

Quality profiles

75.   In the 2008 RAE, producing sub-profiles for outputs worked well, but was problematic for environment and esteem, and there was inconsistency across panels in how this was done. Groups made several suggestions for the REF:

    a.   Some groups suggested a single point rating scale might work better for environment and impact in the REF, and queried whether they need to be combining with outputs into an overall quality profile.

    b.   Some suggested that outputs should provide the core profile, and impact / environment could be used to modify the profile.

    c.   Some were content with the weighted algorithm for combining the elements, except for the 5 per cent rounding.

76.   Some pointed out that that descriptors of 1* to 4* would need to be adapted for environment and esteem.

77.   There was some discussion about the weightings between the elements:

    a.   There was no consensus on the relative importance of each sub-profile other than that outputs should be paramount. One group suggested impact should be between 15 per cent and 25 per cent. Some suggested 5 per cent is too little for any of the elements.

    b.   Some felt that weightings should be imposed centrally with clear and consistent definitions of what goes into which sub-profile. Others felt weightings could be the same for all subjects within a main panel, but vary between main panels.

Annex C

The use of bibliometrics in the REF

Questions

EAG members were asked to consider the following questions about the use of bibliometrics in the REF:

Question 1: To what extent do the pilot citation indicators provide robust and credible outcomes in your discipline?

Question 2: What factors can explain the 'outliers'?

Question 3: Do any of the models have the potential to be applied formulaically in the REF, or should citation indicators be used to inform expert review?

Question 4: Do members agree that the institutional address based model is the least appropriate for use within the REF?

Question 5: What scope is there for the author based model(s) to inform expert review in different disciplines?

Question 6: Would citation information for 'all papers' or 'selected papers' be more useful and informative?

Question 7: Would the benefits and use made of the information outweigh the costs of producing it?

Question 8: What are the priorities for future analysis and further development of indicators and techniques?

Discussion

1.   The following is a full summary of the discussions about bibliometrics in the REF from across all breakout groups.

Robustness of bibliometrics pilot results

2.   In many disciplines (particularly in medicine, biological and physical sciences and psychology), members reported that the 'top 6' model (which looked at the most highly cited papers only) generally produced reasonable results, but with a number of significant discrepancies. In other disciplines (particularly in the social sciences and mathematics) the results were less credible, and in some disciplines (such as health sciences, engineering and computer science) there was a more mixed picture.

3.   Members generally reported that the other two models (which looked at 'all papers') did not generally produce credible results.

4.   A number of reasons for the variations in robustness between disciplines, and for the discrepancies in the results within a discipline, were identified:

a.   Different sets of papers were looked at in the RAE and in the pilot. Some members suggested repeating the citation analysis using only those outputs assessed in the RAE.

b.   The volume of citations and the time taken to accumulate citations varies between disciplines; citation indicators are more robust in disciplines which publish and cite more rapidly.

c.   The coverage of citation databases is limited in a number disciplines, particularly where non-journal outputs are common.

d.   Citations measure impact on the academic community; this is only one aspect of quality, whereas the RAE results represent a rounded view of quality.

e.   Citations do not provide a good measure of applied research and cannot take into account non-academic impact.

f.   Other limitations of bibliometrics that could distort some of the results, such as negative citations.

g.   Some of the sample sizes were small and these tend to be less stable.

h.   More recent papers have had less time to accumulate citations. Even though publication year is taken into account in the analysis, the results were less robust for papers published in the more recent years.

i.   Limitations with the normalisation process. In particular:

i.   The categorisation of journals into fields was felt to be problematic in a number of fields (for example where diverse journals are used, such as in Statistics), and for a number of journals (particularly broad journals that cover several sub-fields such as the Lancet, BMJ, Physical Review, and so on).

ii.   Citation rates were normalised against a worldwide 'mean' for the field; yet the distribution of citations is highly skewed.

j.   Differences in the two commercial citation databases (Web of Science and SCOPUS) led to some marked differences in the results. A few members noted that other databases were more widely used by their disciplines (such as ArchiV and Google Scholar).

k.   The way items are categorised within the databases as 'articles', 'review papers' and so on can differ from the way institutions or researchers would classify them. Some material on the databases (for example in 'trade' journals) would not be considered research.

l.   The mix of sub-fields within a submission can affect citation indicators; for example a submission can be dominated by a highly cited sub-field within Physics.

m.   In a few cases members reported discrepancies between RAE outcomes and citation indicators, where the RAE scores appeared to reflect the prestige of the journals papers were published in, whereas the citation rates for the papers provided a different picture.

Use of bibliometrics in the REF

5.   There was a strong consensus that bibliometrics would be a useful aid to expert review, but that it could not be used formulaically, due to the range of limitations and discrepancies in the data. Expert review would still be required to take these into account and to ensure the credibility of the process.

6.   There was a strong consensus that bibliometrics should be applied to selected papers only. Members agreed that the 'address based' model was undesirable for a number of reasons, not least the substantial problems in associating papers with the relevant UOA. Of the two 'author based' models, members felt that selected papers would be more useful and informative, providing a better discrimination of quality. There was no consensus on what value information on all papers would add, and members raised concerns that assessment of all papers would disincentivise speculative research and lead to other adverse behaviours.

7.   Members discussed a number of ways in which panels could make use of citation data to enhance the reliability and consistency of expert review and/or to reduce panels' workloads. There was no clear consensus on a single approach and members felt that the particular ways in which panels could make use of the data should vary as appropriate to the discipline. The range of possible uses included:

a.   To inform the reading of individual outputs. (Most groups supported this approach although some were concerned about using citation data in this way).

b.   As indicators for each submission as a whole, to sense check or provide a 'challenge' to the panel's scores based on reviewing the outputs.

c.   To inform 'borderline' decisions.

d.   To provide benchmarks against international standards and aid calibration against the quality descriptors.

e.   To inform discussions about consistency between panels, or enable comparisons across disciplines.

f.   To enable panels to sample and reduce the number of outputs to be reviewed in detail. Some suggested that bibliometrics could form part of a stratified random sampling procedure; however some members were sceptical about this and many generally doubted that bibliometrics would enable panels to read fewer outputs.

8.   Members discussed the type of citation information that would be useful to panels:

a.   Many agreed that all panels that make use of citation data should be provided with the same types of data, but that they could use or interpret the data differently as appropriate.

b.   Limitations with the normalisation method were raised. Members generally agreed that panels would want the 'raw' citation count, in addition to data that enables them to interpret this within an international context. This could be a benchmark for the field, or an indication of where the citation count falls within the worldwide distribution for the field (a centile). Panels would also be interested in a benchmark or centile for all papers submitted to the UOA.

c.   There was also interest in the kinds of contextual data provided from the pilot, relating to the sources of citation (local, national and international) and international co-authorship.

d.   There was some discussion about which citation database(s) should be used and many felt that REF should not be limited to using a single database across all panels.

9.   Some issues about the potential behavioural consequences of using bibliometrics were raised:

a.   If bibliometrics were to be used in different ways across sub-panels it could influence institutional decisions about where to submit members of staff, or their decisions about which types of outputs to submit to different panels (for example, selecting on the basis of citations for some panels, and the implication that this could favour older papers or disadvantage early career researchers).

b.   Publication behaviours could be affected, for example a movement towards higher cited journals.

10.   Members discussed whether the benefits of using bibliometrics would outweigh the costs. Some found this difficult to answer given limited knowledge about the costs. Nevertheless there was broad agreement that overall the benefits would outweigh the

costs - assuming a selective approach. For institutions this would involve a similar level of burden to the RAE and any additional cost of using bibliometrics would be largely absorbed by internal management within institutions. For panels, some members felt that bibliometrics might involve additional work (for example in resolving differences between panel judgements and citation scores); others felt that they could be used to increase sampling and reduce panels' workloads.

<u>Further development</u>

11.   A number of areas for further work were suggested:

- Most importantly, to develop normalisation techniques including ways of categorising journals

- Assessing the accuracy of the databases

- Investigating how career stage affects citations (particularly for early career researchers)

- Understanding the behavioural implications of the preferred model

- Investigating other citation sources (such as Google Scholar)

- Understanding coverage of the databases (including, for example foreign language journals)

- Extending the analysis to include conference proceedings.

Annex D

Expert review of outputs

Questions

EAG members were asked to consider the following questions about expert review of outputs in the REF:

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?

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

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

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?

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

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

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?

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)?

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

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

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

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

Discussion

1.   The following is a full summary of the discussions about expert review in the REF from across all breakout groups.

<u>Staff selection</u>

2.   There was widespread agreement that selection of staff by institutions remains the most workable approach. Overall it was acknowledged any approach requiring 'all research active staff' (however defined) to be included, could not avoid institutions re-categorising staff contracts, so in practice few if any of the benefits associated with the universal approach would be achievable.

3.   Members made various suggestions of how the definition of eligible staff should be clarified in the REF:

    a.   There was widespread support for greater clarity regarding fractional appointments, with many suggesting that only those who play a genuine role in the department should be eligible. Some suggested the minimum fractional appointment eligible for return should be increased to 0.25 or 0.5 and that eligible staff should be in post for a set period before the census date. A small number suggested that fractional appointees could be required to submit correspondingly reduced numbers of outputs (e.g. 0.25 FTE = 1 research output), although it was recognised that this could have equal opportunities implications. A small number said the approach to fractional appointments would need to vary across disciplines.

    b.   Many members felt that staff on 'teaching only' contracts should not be eligible. Those on 'teaching and research' or 'research only' contracts should be eligible.

    c.   A few raised concerns about the status of postdoctoral staff, and suggested there should be clear evidence that these are independent researchers, as evidenced, for example, by holding a personal fellowship or being the Principal Investigator on a grant. However, it was acknowledged that this is more difficult for arts and humanities subjects.

4.   The great majority felt overall that there should be transparency about the proportion of staff selected. It was widely agreed that this information could help inform panels' assessment of a department's research environment; and that it would be useful to have this information in order to help interpret the data on, for example, numbers of PhDs awarded per member of staff and to make a judgement about sustainability. However, a number of issues were raised about the interpretation and use of such data:

    a.   Departments could have large numbers of teaching only staff. This was seen as valid, as it releases more time for others to do research. It was considered

important that departments should be given an opportunity to explain their structures and justify the proportion of staff submitted.

b.   Some emphasised that it should be used to provide contextual information rather than used in a punitive way. There was concern that information about all staff could disadvantage small but strong or growing departments and make pockets of excellence less apparent.

c.   There was some concern that individuals on teaching only contracts supporting professional studies programmes would be counted as 'non-submitted' staff, despite them ostensibly being from industry and only teaching one module – this may create a disincentive to use such staff.

d.   Some raised practical concerns about how units that drew staff from a range of Departments / Faculties would define who non-selected staff for any one submission would be.

e.   A small number felt it was not necessary to include information about the proportion of staff submitted. Some were content that small departments / selective submissions in some cases produced high quality results. If HEIs could justify in environment how the research strategy was sustainable (referring to some of the staff not included in the submission) that would be sufficient, but it was not for the panel to judge the merit of the proportion per se.

f.   A very small number felt strongly that information about the proportion of staff submitted should not be included in the submission, as it was up to institutions not REF panel members to form judgements on how research should best be organised within an institution.

Attribution of research outputs

5.   There was widespread agreement that on balance, outputs should be attributed to the author rather than credited to the HEI at the time of publication. A few acknowledged that this approach has implications for the 'transfer market', but generally agreed that this cannot be avoided in any sector or industry and that overall it helps to give the sector dynamism and mobility. This was seen as particularly important for early career researchers.

Category C staff

6.   Most members agreed that research by Category C and D staff should not be included in the REF, except for medical subjects in which staff employed by the NHS make a valuable contribution to the research of the HEI. This would avoid the complexity of establishing whether they had a genuine connection with the department and the numerous audit requests that were involved. It was widely agreed that information about these staff could inform the environment element of the assessment. For example,

panels could be asked to describe the contribution of 'associated' staff as part of the pro-forma on environment.

7.   A number of further points were raised:

    a.   There was some discussion about whether there were disciplines other than medicine in which Category C staff should be retained, and a suggestion that art and design may wish to consider this. A small number suggested the decision on whether or not to include Category C staff should be made by each discipline or at main panel level, with the 'default' presumption that they would not be eligible.

    b.   A small number disagreed that research by Category C and D staff should be excluded. Rather, there should be a tighter definition of Category C staff, requiring a historical contribution to the department and clarity about what is meant by 'close association with the department'. Some of these added that fuller information should be collected about Category C staff in order to minimise audit requests.

    c.   Three groups queried whether Category D staff should ever be included in the assessment, particularly as a retired member of staff is likely to have a replacement that could have been counted in addition. Some suggested Emeritus staff should hold a contract of employment or not be deemed as eligible.

Number of outputs

8.   Members expressed a range of views about the number of outputs to be submitted per researcher and whether panels should have the flexibility to vary this number depending on the discipline.

9.   Some were strongly opposed to any variation of the number of outputs by panel. It was felt that a lack of consistency across panels would make it difficult to differentiate excellence, and could lead to institutions making tactical decisions about which UOAs to submit staff to. Others believed that four outputs is about the right number so it would not be necessary to make variations between panels. Some of these did acknowledge that there may be scope to slightly reduce the number for some disciplines (perhaps two or three outputs), but felt that ultimately this would cause complications around reducing the number of outputs due to special circumstances, such as ECRs.

10.   A minority of members were strongly in favour of panels having flexibility to vary the number of outputs per researcher. For example, this would allow panels to require fewer outputs for subjects where outputs are likely to include monographs.

11.   It was generally agreed that more than four outputs would not be desirable in any discipline, although one group considered up to six might be acceptable in some subjects. Those who felt that the number of outputs could be varied suggested that two should be the minimum.

Double weighting of outputs

12.   Members discussed whether double-weighting certain output types would be a useful way to reduce the burden of assessment. Around half felt that double weighting can be useful and appropriate in some cases and that this could help to reduce the burden of assessment, especially in disciplines where citation indicators are limited or not available.

13.   There was some agreement that where outputs are monograph or book heavy, fewer outputs should be required. Some of these emphasised that the length of the output should be justification for its double-weighting, rather than allowing a long project to be eligible (e.g. significant data collection over a period of time). Many felt that a common definition should be used across panels of what type / length of output would be eligible for double-weighting.

14.   The general view was that if outputs were to be double-weighted, the onus to give them that status should be on the submitting institution rather than the sub-panel (although one group felt strongly that the institution should not make this decision), and that the number of outputs submitted by the researcher concerned should be reduced correspondingly.

15.   Others did not support the idea of double weighting certain outputs, despite acknowledging that double-weighting large books would reduce the sub-panel workload. They considered the approach inequitable. Some added that if this approach was adopted, institutions would be reluctant to take the risk of submitting double-weighted outputs anyway in case they were not assessed as being of high quality.

16.   A few members noted that some submissions in RAE 2008 had included a book as one output and the same book's component chapters as separate outputs, and suggested that this should not be allowed in the REF.

Other ways to reduce panel workloads

17.   Some groups discussed the possibility of sampling submitted outputs. Many felt there were problems associated with sampling. For example, it would be likely to have a 'flattening' effect on quality; there was a risk that whole research groups might be omitted if they were small; it may create perverse incentives such as not to read long foreign language monographs; it may not have credibility with the sector.

18.   Others felt that sampling provided a real opportunity to reduce the burden on panels. There was some discussion about random stratified sampling which could address some of the concerns above e.g. sampling could be done within areas of expertise, by type of output or by seniority of staff. Panels could agree the sampling method in advance and a computer would select outputs to be read within each category. Members would not then be allowed to 'cherry pick' additional outputs beyond the sample selected.

19.   Some noted that random sampling might work for very large submissions, but where there were a number of small submissions it was likely that a majority if not all outputs would still need to be read.

20.   A very small number suggested that journal impact factors could be provided to panels to help them to decide which to read in detail, but there were a number of concerns about this approach.

21.   Some felt that using bibliometrics was the only way to reduce the need to assess all outputs in detail, but raised concerns that in some cases panels may not have confidence in the bibliometrics data to use it in this way in practice. It was felt that bibliometrics should not be imposed on those panels who did not have confidence in them.

<u>Profiling outputs</u>

22.   Some felt strongly that the descriptors of the four star ratings should be reviewed. In particular, members felt that there were too many variants of 'international' in the definitions, that it was difficult to differentiate between 3* and 2*, and that 4* had not turned out to be the pinnacle of excellence originally envisaged. One group felt that the definition of 4* should be clarified as 'in the top 15% of all papers internationally', or something similar.

23.   There was a strong sense that although the definitions might be revised, the categories 1* to 4* should be retained.

24.   Some felt the descriptors must be common to all panels, with no discretion for variation. It was felt that this was important to ensure the fair assessment of interdisciplinary research outputs. A small number felt that panels should have discretion to vary or elaborate them as appropriate.

25.   Members discussed the role of international members, and in general, it was felt that the role of international members had not always been clear and effective in RAE 2008. It was noted that these members were sometimes charged with moderating scores when they had not always been involved in the assessment process.

26.   A significant number suggested that international members should be involved at sub-panel rather than main panel level. There was also a strong feeling that they should be more involved at an early stage of the process such as the initial benchmarking stage. Some added that there could be a role for members in reviewing outputs and again at the end of the exercise to provide some light-touch validation. Members noted that it would require a greater time commitment and remuneration to incentivise international members to take on such a role.

27.   Many suggested that training should be available to international members at the beginning of the process, so they fully understand the context and the scope of their role.

<u>User significance</u>

28.   Members discussed whether, in the REF, statements about user significance should be provided for outputs with particular significance to users. Some commented that in RAE 2008, statements about user significance had been helpful for some disciplines (e.g. Engineering and Computer Science) and that these panels would wish to invite these statements again.

29.   It was agreed that there should be clear guidance about the content of these statements where used, and that lengthy and unverifiable descriptions should be avoided. It should be clear who the users are and how the outputs were useful to them.

30.   Some subject representatives, especially in arts and humanities, felt that it would be difficult to provide such statements for their subjects. There was also widespread concern that the introduction of the impact element of the framework already gave subjects an opportunity to demonstrate significance to users, and that there is a danger this would be double-counted if it also considered under outputs. This needs to be given careful thought in order not to disadvantage certain subjects. Indeed, some felt that it was not possible to consider the place of statements about user significance until arrangements for the impact element of the REF were more developed.

31.   A small number felt strongly that if there is an opportunity to indicate the significance of an output to users in this way, there must also be an opportunity to indicate significance to the academic community.

Annex E

Assessing research environment in the REF

Questions

EAG members were asked to consider the following questions about assessing research environment in the REF:

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

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

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

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

Question 5. What other key indicators should be included (as part of a common family of indicators for research environment)?

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

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

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

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

Discussion

1.   The following is a full summary of the discussions about assessing research environment in the REF from across all breakout groups.

Scope and focus of research environment

2.   Members discussed which key elements the environment element should focus on. Many commented that the elements should be similar to the RAE 2008. In particular, sustainability, development of staff, and the research strategy were important.

3.   Some added that the research strategy should be a forward looking document which includes explicit deliverables, rather than a general overview. A few suggested that it would be helpful to make comparisons with a department's previous strategy.

4.   Some groups also felt the following were important aspects of the environment element: enabling behaviour (such as impact and public engagement), collaboration, grant income, some aspects of 'esteem' that contribute to sustaining the discipline (such as editorship of publications), although others felt this should be captured within the 'impact' element. Additional suggestions included details about non-submitted staff and details about facilities.

5.   There was some debate about whether the wider environment or whole department should be assessed, rather than focusing on selected staff:

    a.   Most members felt that information about what non-submitted staff added to the department would provide useful context. Others thought this went beyond the RAE / REF remit to judge research quality and would make the process more burdensome, so the focus should be on information available for selected staff only.

    b.   Some commented that the best approach would be to combine an assessment of the whole department with relevant information about selected staff (such as income and investment) where available.

    c.   Overall it was felt that data about the profile of all eligible staff in the department would be valuable to the assessment of environment, providing 'eligible' and 'department' could be adequately defined.

Indicators and narrative

6.   Members discussed whether data about PGRs and research income that is available from HESA would be sufficient for the purposes of the REF:

    a.   It was felt that taking data on PGRs and research income from HESA would help reduce burden, but there were some doubts about the practicality of this at present. There were concerns that HESA data may not be reliable, does not map onto UOAs, and would not provide information about the source of funding for PGR students.

    b.   Some felt that HESA data would be sufficient for the purposes of the REF, but that care must be taken to include measurements of department growth or contraction over the assessment period.

    c.   Members felt that standard quantitative analyses should be provided to panels to inform their assessment of environment in the REF. In particular,

members mentioned the following: numbers of students undertaking higher research degrees, numbers of higher research degrees awarded, numbers of studentships and fellowships, investment in capacity building, support for the development of career pathways for young researchers, destinations of students after award of PhD, research income from public / charitable sources and from industry, national and international research collaborations, participation at or hosting of conferences.

d. Some members considered that research Masters should be excluded from PGR numbers, as they were more akin to a taught qualification, but others disagreed. Some groups also discussed whether information about number of research student publications or prizes was meaningful.

7. It was widely agreed that the RA5 or equivalent should be a standard pro forma with specific questions and word limits. Several issues were raised about this:

a. Some considered whether a smaller department should be required to submit a shorter statement, but did not come to a conclusion about this.

b. Some groups considered what key questions about the research environment should be included in the pro forma. Some suggested that there should be an opportunity for departments to demonstrate the vision and management of a unit and illustrate sustainability and forward-planning. Many emphasised that all evidence submitted should be verifiable.

c. Some suggested that sub-panels should have the flexibility to tailor the pro forma for their specific disciplines. Others suggested that the questions should remain fixed across all panels (partly to make it easier for HEIs to complete), but that sub-panels should be able to vary the relative weightings of each questions.

Granularity and criteria for assessment

8. Members discussed whether research environment should be assessed using a single point scale rather than a profile, and expressed a mixture of views:

a. The majority were in favour of a profile. Some suggested it would be possible to have a number of elements within environment which could be assessed individually and combined into a profile. These elements and their weightings would be transparent and consulted on in advance.

b. Some felt strongly a single point scale should not be used. There were concerns that a single-score would reintroduce the 'cliff edge' effect that the profile was designed to mitigate against. There was also a strong feeling that the result should be an overall profile, rather than three sets of scores which may be open to funding manipulation by government.

     c.   Those in favour of a single score thought the heterogeneity of some sub-panels made it difficult to consistently judge environment at such a detailed level. Some commented that those who had taken a formulaic approach to the assessment environment in the RAE 2008 regretted this approach, as there were some unexpected returns which skewed results.

9.   Some members considered what the criteria and descriptors should be for the assessment of environment. Many felt the descriptors should be reviewed and commented that the descriptors should be different from those for research outputs, since terms such as rigour and originality are not applicable to the environment (though a small number disagreed with this). Some suggested that single word descriptors would be suitable, e.g. outstanding, excellent, satisfactory, unsatisfactory. Others suggested ratings such as 'likely to result in the increase of positive research activity' or 'likely to result in a continuation of the current level of research activity', etc.

10.   Members considered whether we should consider assessing environment at a broader (main panel) level. There was widespread agreement that environment should continue to be assessed at sub-panel level, and not by main panels. Some commented that the main panels are too heterogeneous to undertake this task. The assessment should take place at sub-panel level in order to make appropriate judgements about facilities and income etc. Some suggested that where there is a mismatch between the quality of outputs and environment, it is necessary to establish the reasons for this, so both assessments should be undertaken by the same group of people.

Annex F

Taking account of impact in the REF

Questions

EAG members were asked to consider the following questions about taking account of impact in the REF:

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.

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

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

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

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

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

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

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

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

Discussion

1.   The following is a full summary of the discussions about taking account of impact in the REF from across all breakout groups.

General approach

2.   Overall, most members agreed or accepted that the impact of research should be assessed as part of the REF. Most agreed that the REF should take a broad overview of the breadth and depth of impacts achieved by a department.

3.   There was agreement that the assessment should focus on a limited number of selected examples rather than try to be comprehensive or attempt to link all outputs to impacts.

4.   Members agreed that the definition of impact should be broad and should include benefits to the economy, society, public policy and services, culture, health, wellbeing and quality of life. There was a strong and widespread view that academic impact should also be included. Examples were given of subjects such as mathematics which may have minimal direct impact on the outside world themselves, but which heavily influence subjects such as engineering, which in turn have a direct impact. This academic impact should be seen to be just as valid as more direct impact on the economy or society. This was emphasised by some user representatives as well as academics.

5.   Some other concerns were raised:

a.   Some felt that outputs which do not have an obvious impact could be disadvantaged and that by focusing so much on impact, researchers would be discouraged from basic or blue skies research.

b.   A very small number felt that impact was difficult to define and quantify in some disciplines, and suggested impact was not really relevant to their subjects.

c.   A minority were concerned that the proposed approach places too much emphasis on impact, in that user-significance and engagement is already measured in the outputs and environment elements. Some suggested therefore that impact need not be a separate element the REF, and could perhaps be combined with environment. These members felt that if impact was to be a separate element, it should be assigned a low weighting.

d.   Others suggested that the weighting for impact should be as little as 5%, especially for the first REF exercise when the approach will not be tested and fully understood. However, some acknowledged that HEFCE is unlikely to accept a weighting below 10% for impact.

6.   Some groups offered suggested weightings, for example:

Minimum of 75% outputs and a maximum of 25% for impact and environment
Minimum 70% outputs and a maximum of 10% for impact
Set weightings of 60% outputs, 20% impact and 20% environment.

Evidence of impact

7.   There was a general consensus that impact should be assessed through a combination of narrative evidence, indicators and case studies.

8.   Some agreed that the narrative statement should be presented using a structured pro forma with specific questions. Some suggested there should be a strict word limit to keep the assessment manageable. The statement should address the departmental strategy and how it fosters impact, identify the kinds of impact achieved (perhaps selected from a basket of possible options) and include information about users who had

benefited from the research (who could be contacted to verify the case studies – perhaps through a random sampling approach).

9.   Many stated that it would be important to limit the number of case studies in order to keep the assessment process manageable. Some suggested between three and six case studies per submission would be appropriate. Some suggested the precise number might vary depending on the size of the department or the number of staff submitted. There should also be a word limit for each case study.

10.   Many expressed concern that interpreting the narrative statement could significantly add to panels' workload. Departments should be discouraged from simply giving website links and expecting panels to search for the evidence of impact themselves. Evidence should not be too broad and generalised.

11.   A few commented that the case-study approach would be a good opportunity to 'un-pack' some interdisciplinary examples.

12.   Many members stated that sub-panels would need to decide what indicators of impact should be relevant for their subject. As a starting point, some members suggested considering the research indicators listed in the draft report by RAND Europe.

13.   Some discussed whether suitable indicators exist for arts and humanities. Some debated whether some of the following would be appropriate: a research monograph re-issued as a text book; a crossover book which made research accessible in a popular format; a popular history series on television. Some highlighted concerns about where to draw the line between research and knowledge transfer.

14.   A significant number of members felt strongly that 'public engagement' should qualify as an indicator of impact. Some added that evidence of 'impact enablers' – i.e. behaviours that are likely to produce impact in the longer term (such as engagement with industry or users), were also important. HEIs should be encouraged to demonstrate the mechanisms they have in place to generate impact and the infrastructure to facilitate translational activity.

15.   A small number suggested that these 'impact enablers' could be used as evidence during a transitional period, until HEIs are more practised at collecting and demonstrating other evidence, or until other indicators of impact are more robust and accepted. A few members suggested that over time, HEIs would build up experience in demonstrating impact as this becomes a more routine requirement across the sector.

16.   A small number of members felt strongly that definition of impact and the general approach should be the aligned with that of the Research Councils.

Time lags and attribution

17.   A number of members expressed support for the approach outlined in the EAG paper 2/2 – i.e. that the REF should look at impacts that have become evident during the assessment period, arising from research undertaken over a longer time frame. Some suggested there should be guidance and a time-limit on how far back one could go. In some disciplines (e.g. neuroscience and cardiovascular medicine) there may be time lags somewhere between 15-25 years, but it was felt this would be too long to assess in the REF. Some noted that this would be easiest to do in cases where researchers at the institution were still working on related areas.

18.   Some expressed concern about attribution and felt that substantiating claims would be impossible in some cases. A few also raised the issue of restructuring in HEIs and whether a department should be able to claim credit for the work of a predecessor department.

19.   Some explored how the issues of time time-lags and attribution are linked to the question of what is being assessed: the HEI, the department, or the impact of the individuals submitted. Some did not agree that the impact of a department should be assessed. A few felt that if impact arises from a specific piece of research, then the author should perhaps be able to take it with them. There was a suggestion that older impacts could stay with the department, but newer impacts could be portable, and that a submission could demonstrate either or both of these.

20.   Members felt that several institutions should be able to claim credit for impact from collaborative research projects. This would parallel the existing approach to collaborative research publications.

Criteria and rating scales for impact

21.   Members discussed what the criteria and descriptors for impact should be. Some suggested there should be single word descriptors for impact e.g. outstanding, excellent, satisfactory, unsatisfactory, as these would be the least ambiguous. Some commented that 'profound' was not a useful word to use in the descriptor.
22.   Some suggested that panels should be able to vary the descriptors to fit their disciplines.

23.   Some emphasised that the criteria should include intellectual impact, although a number of users expressed concern that this might have the effect of double-counting academic impact.

24.   Members expressed mixed views about whether impact should be assessed using a single point scale or a profile, but overall, members preferred the option of a profile. Those who favoured using a profile were concerned that the effect of a single point rating on the overall quality profile could distort the overall results. If a profile was used, weightings could be applied to different elements and different case studies to allow panels to create a profile.

25.   However, it was also suggested that it might be possible to combine a single point rating for impact with the quality profile for outputs by using a multiplier across all elements of the profile, producing a weighted distribution.

26.   Those who supported a single point scale felt that a fine grained assessment would be too difficult and artificial for this element.

27.   There was some discussion about the level of subject granularity at which impact should be assessed. There was general agreement that impact should be assessed at the sub-panel level. Some added that granularity could vary by sub-panel and could be by individuals, research groups or the whole submission.

Research users

28.   Many felt that users should have a key role in validating case-studies in the REF. For example, research users' details could be supplied on case studies to and a sample could be contacted to corroborate impact claims. (However, there were some concerns from social scientists about getting feedback from government departments when research had provided 'unwelcome' views).

29.   Others suggested that users should be consulted on the design and criteria setting, but may not need to be full members of sub-panels. In particular, users should not be expected to read large volumes of outputs; their role on panels could more usefully focus on assessing the impact element.

30.   Many suggested that it would be difficult to engage sufficient numbers of users in the REF, especially as the time commitment is likely to be substantial. It was suggested that HEFCE begin by holding a workshop for users (including those involved in the RAE and others) to discuss the best way to involve them in the process.

Annex G

Equalities issues

1.   The following is a full summary of the discussions about taking account of impact in the REF from across all breakout groups.

2.   In general, members agreed the guidance and training for panels and the guidance given to HEIs was sufficient in RAE 2008, but that more could be done to ensure greater consistency and clarity in the REF.

3.   Many discussed the difficulties in assessing individual staff circumstances, especially when HEIs provided a lack of explanation. The vast majority supported the idea of a central team (with support from HEFCE) to develop consistent criteria and guidance for all UOA, and potentially to adjudicate claims to special circumstances, or to provide guidance to panels on this. Although this involves an additional group in the process, it was felt that such a body would be well placed to achieve more consistent treatment of special circumstances, and relieve panels of some of their workload. Some emphasised that it would need to include some academic representatives.

4.   Many felt there had not been clear enough guidance on the treatment of ECRs in RAE 2008. Overall, it was felt there should be a common definition and common treatment of ECRs across the exercise. Many felt that the minimum number of outputs expected from an ECR should simply be pro-rated e.g. X years = 1 output. There was an acknowledgement that it can be difficult to define ECRs and that comparing them in this way could be problematic, but it was felt important to tighten up the definition from 2008. There was some agreement that if ECRs submitted more outputs than required, all outputs should be considered.

5.   Many members were concerned about fractional appointees being able to submit the full number of outputs, and some felt there should be strict pro-rata rules. However, in some disciplines where the majority of staff were on such fractional appointments, this could cause problems.

Annex H

The three elements of assessment

Questions

EAG members were asked to consider the following questions about combining the three elements of assessment:

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

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

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

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

Discussion

1.   The following is a full summary of the discussions about combining the three elements of assessment in the REF from across all breakout groups.

Assessing impact

2.   Members discussed the proposed approach to impact and whether it is workable across all disciplines. There was general support for the broad approach proposed for assessing impact, although a number of issues were raised.

3.   A significant number raised concerns about the following:

- Using a narrative approach to assessing impact will be burdensome for panels and encourage 'creative writing' by HEIs, similar to the RA5s in RAE 2008. There would also be considerable burden for HEIs in producing the case studies, finding evidence, and checking references.

- There is a danger impact could be double-counted through assessment of both the significance of outputs and of impact as a separate element.

4.   A smaller number raised the following issues:

- The suggestion that an excellent department should produce a portfolio 'including work that is world-leading', is likely to favour larger departments.

- The definition of impact should explicitly include cultural and quality of life impacts, rather than just impacts to the economy and society.

- How would impact be measured if it occurred in other countries?

- A small number expressed a preference for the term 'beneficiaries' over 'users'.

- Public outreach / engagement should be included within the impact element.

5.   Most groups discussed how academic impact should fit into the assessment. Around half expressed support for the suggested approach to academic impact, agreeing that academic impact would be covered sufficiently within the outputs and environment elements, and should not also be reflected in the impact element. Many of these were reassured by the suggestion that under the impact element, impacts on other disciplines could be included, for example where a discipline such as pure mathematics influences other disciplines which in turn have impact on the economy or society.

6.   However, a small number felt strongly that academic impact should be included within the impacts element, including impact within a discipline. Some argued that there can be a long chain within and across disciplines before impact is realised.

7.   Others felt that academic impact should not be included at all within the impact element: it would be difficult to substantiate impact on another discipline and it is already covered within the outputs element.

8.   There was discussion in some groups around whether impact should relate to the institution or to the researcher. Overall it was agreed that the research activity should be attributable to the institution, although it could be the work of one researcher or a team of changing constitution over time.

9.   Many emphasised the importance of HEFCE producing clear, timely guidance on the assessment of impact for both panels and HEIs.

10.   There was widespread support for testing the approach to impact in advance of the REF through an impact pilot.

Weightings between the three elements

11.   There was general agreement that there should be standard weightings between the three elements (outputs, environment and impact) across all disciplines.

12.   However, many agreed that while there should certainly be much less flexibility between weightings than in RAE 2008, some exceptions may be necessary. It was suggested that any deviations from the default weightings would need to be justified and presented clearly.

13.   The great majority agreed that outputs should be at least 60%, and some felt that 70-75% would be about right. For some subjects, 60% for outputs would represent a significant reduction from RAE 2008. One group felt that 50% would be adequate for outputs.

14.   Many members were concerned about having a weighting of 20% or more for impact, especially for the first round of the REF and given that assessment in this area is necessarily subjective. Nevertheless some of these accepted that any less than 20% would be unacceptable politically. A small number, especially in the medical groups, felt that 20% or more for impact was about right.

15.   Overall the suggested weightings fell between these ranges:
Outputs: 60%-75%
Environment: 10%-20%
Impact 10%-20%

Profiles or single point scales for environment and impact

16.   Members discussed whether impact and environment should produce sub-profiles or single point ratings, and how these should be combined into an overall quality rating.

17.   The majority of members felt that both impact and environment should produce profiles. The profiles could be made up of scores given for different aspects of environment / impact. Some suggested these aspects should be consistent across main or sub panels where possible and made public in advance.

18.   Many emphasised the importance of having clear level descriptors for environment and impact, which are distinct from the 'outputs' level descriptors, and made available to the HEIs in advance.

19.   There was widespread agreement that there should be clear guidance on profiles and how they should be constructed by panels, in order that there is greater consistency in the way the profiling is done across the exercise.

20.   A small number discussed whether the sub-profiles should be published at the same time as overall profiles, and saw no reason why this shouldn't be the case (although in RAE the sub-profiles were released separately from the results).

21.   Some discussed the option of combining single point ratings for impact and environment with an outputs profile by using the ratings as multiplying factors, but only one group was in favour of this approach. Most felt that this would not be workable in practice.

Annex I

REF Panels

Questions

EAG members were asked to consider the following questions about 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?

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

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

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

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?

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

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

Discussion

1.   The following is a full summary of the discussions about REF panels from across all breakout groups.

Units of Assessment

2.   On the whole, members agreed that there should be fewer UOAs in the REF, with fewer fluid boundaries between them and less variation in the volume of work covered by each panel.

3.   Some members stated that in the RAE, some panels covered diverse fields, such as the Agriculture, Veterinary and Food Science panel; and that this had been workable in practice. The important issue is whether or not there is enough expertise available within the panel to make sound judgements.

4.   Members identified various sub-panels which could be combined or merged into larger UOAs in the REF, such as:

- Pure Maths, Applied Maths, Statistics and Operational Research
- The UOAs covered respectively by main panels A, B and C
- Biological Sciences and Pre-clinical and Human Biological Sciences
- Earth Systems and Environmental Sciences and Physics

5.   However, a small number expressed uncertainty about HEFCE's motivations for reducing the number of sub-panels and they could see no clear justification for this. Some felt that the number of FTE submitted to RAE 2008 was not in itself a clear indicator of the workload of sub-panels. In general it was felt that consistency rather than workload should be the main driver for merging UOAs.

6.   Some encouraged HEFCE to consider what impact changing the panel structures would have on HEIs' structures.

7.   They also discussed the suggestion that greater use could be made of reviewers who are not full members of panels, to share the workload and ensure the full range of sub-fields are covered. There was some support for this approach, but others expressed concern that the legitimacy of the system could be seen to be undermined if there is less full-time expertise on the panels.

8.   Some expressed concerns about outputs being reviewed by those who were not fully involved in the assessment process and felt it might be difficult to formulate a 'panel view'. Appropriate mechanisms would be needed to calibrate and ensure sufficient consistency of marking.

9.   Some cautioned that the nominations process for experts would need to less time-consuming and cumbersome than in the RAE 2008, when a number of specialist advisers were appointed late in the process. Some suggested the chair of the panel should appoint these specialists, without first seeking nominations from other bodies.

Consistency

10.   There was general support for greater consistency across the exercise as a whole.

11.   The great majority of members agreed with the suggestion in the paper that key areas should be consistent across all UOAs.[1] (However, one group raised objections to the suggestion of standardising the maximum number of outputs submitted per researcher. It felt that consistency in this area is meaningless as the number still

---

[1] These areas included the following: the assessment of outputs, environment and impact against generic criteria and quality definitions; the weightings between the three elements; the definitions of research and eligible outputs; the definitions of staff eligibility and criteria for special circumstances.

represents a different volume of work. Another groups suggested bigger panels may wish to require fewer outputs.)

12.   Some groups also discussed options for reducing the maximum number of outputs. Several groups favoured a maximum of three outputs per person; a small number felt that two would be adequate; and some others felt that four outputs should be retained.

13.   The groups discussed how the REF can support consistent application of internationally benchmarked standards across the panels. Members agreed that while it would be unrealistic to expect outcomes to be directly comparable across all UOAs, there should be consistent application of internationally benchmarked standards.

14.   Some discussed how bibliometrics could be used for international benchmarking in some disciplines and many supported the use of bibliometrics where available.

15.   International members should also be used, and some comments were made about how the role of international members could be strengthened in the REF:

- International members could be involved at sub-panel rather than main panel level (particularly if there is a broader main panel structure). There were mixed views about whether or not international members should attend all of the sub-panel meetings.

- There should be clear guidance and procedures in advance on the role of international members across the framework, rather than these being negotiated on a panel-by-panel basis.

- Thought should be given to the recruitment process for international members, and those who are appointed should have sufficient time to contribute helpfully.

16.   A few also suggested the REF could make use of research council international review information.

Main panels

17.   There was strong support for broader main panels. Most members felt that having fewer main panels would help to achieve greater consistency. Several groups supported having between four and six main panels.

18.   Several points were raised about the way the main panels should operate:

- Some suggested that representatives from main panels should meet early in the process to agree areas of consistency.
- Some suggested cross-membership of main panels for UOAs that straddle boundaries between them.

- In RAE 2008, panel advisors were helpful in achieving consistency and their role should be retained and enhanced to ensure consistency in a new structure.
- Sufficient time should be built in at the end of the process to ensure the possibility of adjustments to assessment across panels.

19. There was broad agreement that the areas outlined in the paper should be consistent within main panels wherever possible.[2]

20. One group felt that 'Approaches to considering other relevant details' about outputs including statements of user significance or descriptions of an author's contribution to multi-author outputs' did not need to be the same between panels as it varies between disciplines.

Nominations processes and membership

21. Members discussed nominations processes and panel membership.

22. Many agreed there were some difficulties with the process for nominating and recruiting specialist advisers in the RAE, especially within the short timeframes, and that this would need to be reviewed. This is particularly important if in the REF more members are likely to have a specialist role.

23. Many agreed that time should be built in to train international panel members, user members and specialist advisers before the assessment process begins.

24. Some noted issues with multiple members on panels from the same institution. This meant that large parts of the panel had to withdraw because of conflicts of interest. However, it was noted that sometimes this is inevitable.

25. Some felt that HEFCE needs to do more to promote the importance of the role of users on panels to encourage more to take part, and define the role clearly in advance. There was a suggestion that more users are likely to be interested in getting involved in the process if the time commitment is less onerous than in RAE 2008. There was some agreement that there should be greater representation of users on sub panels but that their roles should be focussed on impact and environment rather than reading outputs.

---

[2] These areas included the following: similarity of criteria and working methods; the main components to be considered within impact and environment, and the relative emphasis between these components; comparability of outcomes.