

October 2006

# Use of research metrics in the arts and humanities

**Report of the Expert Group set up jointly by the Arts  
and Humanities Research Council and the Higher  
Education Funding Council for England**

# Use of research metrics in the arts and humanities

## Executive summary

1. In June 2006, the Chief Executives of the Arts and Humanities Research Council (AHRC) and the Higher Education Funding Council for England (HEFCE) jointly established an expert group to explore the use of metrics in the assessment of arts and humanities research. This report outlines the group's recommendations. It has three sections: an introduction, key principles, and operational features.

2. The introduction gives a brief account of the background to the group's work, its deliberations and the consultation that it undertook.

3. The report then outlines a number of key principles which should inform any changes to the assessment of research that would impact on arts and humanities disciplines. These principles hold that:

- the dual support system for research funding should be maintained
- for the purposes of research assessment, there is no fundamental difference between STEM subjects (science, technology, engineering and mathematics) and arts and humanities disciplines
- nevertheless, an assessment system needs to be sensitive to disciplinary differences
- the assessment system should reflect an evolving research landscape
- peer review of outputs should be maintained, although steps should be taken to reduce the assessment burden on reviewers
- only a portfolio of metrics can reflect the totality of the research process
- the assessment system should encourage the sustainability and vitality of research environments
- equal opportunities issues should be given due consideration when adopting any new research quality assessment system
- the assessment system should be sensitive to the possibility of generating perverse incentives that promote unhelpful research or management practices
- it is important to develop robust metrics for the purposes of research management
- a funding cycle of 5-7 years should be retained.
- the UK should continue its work in developing methods for assessing arts and humanities research, and seek ultimately to secure agreement on the standards and methods to be used in international benchmarking exercises.

4. The section on operational features considers how the assessment might be implemented, bearing in mind the principles outlined above. It contains a specific proposal for lightening the assessment burden on peer reviewers, using a method of sampling of submitted outputs. It also proposes other reforms to the system of peer review.

5. Finally, the report details how both peer review and a range of metrics might be brought together as elements of an overarching framework. These elements are listed in paragraphs 18-29, with a number of considerations that should be kept in mind when deploying them as part of a research assessment system for the arts and humanities.

## **Introduction**

6. In June 2006, the Chief Executives of AHRC and HEFCE jointly established an expert group to explore the use of metrics in the assessment of arts and humanities research. The terms of reference and membership of the group are outlined at Annex A. The group was established in response to the consultation on reform of research assessment overseen by the Department for Education and Skills (DfES), in order to provide specialised input from the perspective of the arts and humanities research community.
7. Three meetings of the group were convened between July and October 2006. Issues discussed at those meetings included the use of metrics in Research Assessment Exercises (RAEs) in 2001 and 2008, the responses arising from the consultation with the arts and humanities community, and equal opportunities in research assessment. In addition, the group's deliberations were informed by background information that provided much of the context for debate and discussion. The full list of information made available to the group is at Annex B.
8. The group's membership was designed to reflect the diversity of the arts and humanities subject domain. However, it quickly became apparent that extensive consultation with the wider academic community would be necessary in order properly to reflect the very different disciplinary perspectives on research assessment. In September 2006, the group embarked on a wide-ranging consultation exercise with major groups of academic stakeholders including RAE panel chairs, AHRC panellists and panel chairs, members of the AHRC Council, subject associations and HEIs. The views elicited underpin many of the key principles and proposed operational features of the assessment framework for research in the arts and humanities that are outlined in this document.
9. The term 'peer review' is used frequently in this report; it also occurred repeatedly in all the group's discussions and consultations. It should be understood here as referring to judgements made by both academic and non-academic experts where appropriate (see paragraphs 18-23 below). Analogously, the term 'peers' should be understood throughout the report to mean both academic and non-academic experts.

## **Key principles**

10. The group believes that the following principles should inform any changes to the assessment of the research performance of UK HEIs:
  - (i) The system of dual support funding for research in HEIs, through the distinct functioning of its two elements (the funding councils and the Research Councils), is important for the long-term sustainability of a world-class research culture in the UK. It should be maintained. Not only does this guarantee the integrity and autonomy of research conducted in HEIs, but

there is a crucial interface between the funding councils' quality-related (QR) funding and Research Council funding streams, in that many of the excellent projects supported by Research Councils have been developed using QR funding. The success of Research Councils in supporting excellent research is dependent on the health of the dual support system.

- (ii) Principle (i) is particularly apt in the case of the arts and humanities, where the QR stream represents the overwhelming majority of the funds available for research. In proposing changes to the system of research assessment, it should be recognised that perturbations in QR funding may create instability that may disproportionately affect the capacity of arts and humanities researchers to produce world-class research.
- (iii) There is no fundamental difference in the nature of the research enterprise in the science, technology, engineering and mathematics (STEM) disciplines on the one hand, and the arts and humanities on the other. Rather, these disciplines represent a continuum of research endeavour, along which methods and resource requirements vary in ways that do not map easily onto the current subject divisions. The demand for research inputs ranges along the spectrum from resource-intensive disciplines, such as chemistry and archaeology, to non-resource-intensive disciplines such as mathematics and philosophy. The disciplines that make up the arts and humanities are distinctive in their approaches and concerns but should not be considered exceptional.
- (iv) This distinctiveness is also apparent at the level of discipline. Although it should be possible to devise a broad framework of assessment that applies to all disciplines, the nature and scope of the elements of that framework should be sensitive to the distinctive characteristics of each discipline. These include the size of the research community, its demand for inputs, the inputs available to it, its publication patterns, and the nature and organisation of the research process.
- (v) It is recognised that the sensitivity to disciplinary distinctiveness and the varieties of research called for in (iv), when embedded in an assessment system linked to funding, may have an impact on interdisciplinary behaviour. It is important that any proposed system does not inhibit interdisciplinary collaboration.
- (vi) It is also recognised that the metrics most appropriate to curiosity-based research, to practice-based research and to user-focused research, may have distinctive features. Researchers in some disciplines in the arts and humanities (for example, the creative and performing arts and linguistics) are involved in practice-based and user-focused research that is closer, by way of experimental method, to some of the research in STEM subjects than it is, for example, to archival research undertaken in history.
- (vii) The research landscape is evolving in ways in which it was impossible to anticipate when the RAE was established 20 years ago. The growth in collaborative and interdisciplinary research, the increasing use of information technologies in all facets of the production and organisation of research, the emphasis on the wider dissemination of its outputs and outcomes, and the

- embedding of the creative and performing arts in the landscape have meant that a more holistic approach to research assessment is now required.
- (viii) This holistic approach will involve the combination of a number of elements to gain a more rounded picture of research quality and performance. There is a danger that, in focusing solely on the quality of outputs, valuable research activities and collaborations that have developed in the past 20 years are neglected or discouraged, since they are not well served by that focus. It will be important to redress the balance between inputs, activity and outputs in order to give more recognition to the elements that sustain world-class research cultures and ensure their wider relevance.
- (ix) While both the dynamic nature of the research landscape and our knowledge of the best measures to assess research quality are recognised, research outputs are currently the most reliable indicator of such quality. With time, experience and research, credible quantitative methods could emerge. At this point, however, it will be necessary to retain the application of human judgement through peer review processes as an element in the overall assessment framework.
- (x) Peer review processes specific to the assessment of research quality should be retained for the foreseeable future, but it is necessary to relieve the assessment burden on reviewers because this burden has become unsustainable. This must be done in a way that does not weaken confidence in the process and preferably enhances the effectiveness of the peer review process as a whole.
- (xi) There is no single metric that is appropriate to measuring the performance of arts and humanities research. Of the metrics available, some are well established, while others are being developed. For example, in areas where outputs may be non-textual, such as in the creative and performing arts, the most appropriate and reliable measures are still being examined and refined. Nonetheless, it should be acknowledged that the UK is in advance of other nations in seeking to identify the best measures for 'non-traditional' forms of research output.
- (xii) The metrics chosen as part of the assessment framework should reflect the multiplicity of peer review systems which are already in place and are an integral feature of academic life. They include peer review of books, journal articles, exhibitions, performances and other outputs; and peer review of applications for project-based and infrastructure grants. Future research assessment systems should aim to integrate these current practices and other peer review systems, where robust and transparent, in an effort to lighten the current assessment burden.
- (xiii) The assessment framework should be designed to capture the totality of the research process of a unit (whether discipline-based or interdisciplinary) in which the whole may be greater (or less) than the achievements of individual researchers. In attempting to measure the totality of the research process (from inputs, to activity, to outputs, to outcomes), it is necessary to consider a balance of metrics and judgements which allows a profile of research performance to emerge.

- (xiv) With respect to metrics, attention should be paid to the quality of the current data-sets, recognising that some metrics are currently more robust and auditable than others, and that they may be applied to some disciplines more readily than others. For example, the clear deficiencies of commercial citation indices in terms of their coverage of arts and humanities outputs make the use of bibliometric indicators for assessment purposes highly problematic at present. This is despite promising developments in public-access bibliometric tools and publicly-funded initiatives both in the UK and internationally. However, as bibliometric science develops and if the data become more sophisticated, developments in these and other metrics designed to measure the significance of outputs should be monitored and their 'fitness for purpose' reassessed.
- (xv) It is understood that the dynamism, vitality and sustainability of a department or research group depend on factors such as the willingness to recruit early-career researchers. Such factors are not easily measured by standard indicators of input, activity and output. An assessment system linked to funding has the potential to reward such positive contributions to the research environment and other desirable behaviours, either directly through appropriate metrics, or through an auditable contextual narrative provided by the submitting department or group.
- (xvi) Due consideration must be given to equal opportunities in the establishment of any research assessment system.
- (xvii) It is important to avoid, or at least minimise, possible perverse incentives and negative behaviours in the establishment of any research assessment system.
- (xviii) In addition to their use in the assessment and funding of research, the benefits should be recognised of developing a range of metrics which provide timely and accurate information for institutional managers on all aspects of research activity.
- (xix) To ensure stability of institutional planning, a funding cycle of 5-7 years is recommended. This would facilitate orderly planning within institutions, while still allowing the research base to respond to changing circumstances. In order to avoid large-scale data-collection exercises once every 5-7 years, it would be prudent to require institutions to collect data annually, as a necessary part of their normal work in producing high quality research and as an indispensable tool in assisting research managers to monitor and plan the production of that research.
- (xx) Furthermore, the aim of producing world-class research cultures should be properly evidenced through a system of international benchmarking. The UK already has an internationally leading role in the development of methods for assessing arts and humanities research. The development of accurate and workable metrics that reflected the distinctiveness of research process in those disciplines would enhance that role further, facilitating agreement on the standards and methods to be used in international benchmarking exercises.

- (xxi) The UK's position as an international leader both in the production of research and in the processes of research assessment is the result of significant public investment made over the past 20 years. The recommendations that follow seek to protect and build upon this public investment through refining and advancing the knowledge and expertise developed within the UK research community in producing world-class research and the processes to evaluate it.

## **Operational features**

11. The preceding principles constitute the broad framework for a system of assessment that reflects the dynamic landscape of research in the arts and humanities and ensures its stability and intellectual dynamism into the future. The following paragraphs give further details of what the group believes to be the operational features necessary for a fair, robust and credible system. The key features are as follows:

- the units of research assessed should be departments or groups (not outputs or individuals) and the assessment should be of their research endeavour in all its aspects and dimensions
- the assessment framework should incorporate a range of interrelated metrics that, in combination, will be an efficient and robust component of a system to assess research quality
- the metrical elements should each be weighted within the overall framework to balance, prioritise and incentivise the whole research endeavour
- operationally, the assessment framework should be delivered through significantly lighter-touch peer review and greater use of metrical data
- the metrics identified should facilitate institutional planning of high quality research.

12. Whatever the precise nature of the metrics chosen and their individual weightings, it is essential that a trial of their use be compared with the results of RAE 2008, so that the efficacy of a metrics-informed process can be evaluated. This will be vital in gaining the confidence of the arts and humanities community. Comparing the results of RAE 2001 with the metric data available at the time would allow further analysis and testing of the utility of metrics, in recognition of the fact that the use of metrics in research assessment is a continuous learning process.

13. In order to reduce the assessment burden, the assessment system should consider many fewer and much broader reporting units than the units of assessment (UoAs) currently in place. For example, the individual UoAs for French, Italian, German Studies and so on could be merged as one Modern Languages UoA, and similar broad groupings of cognate and complementary disciplines should be considered elsewhere.

### *Overall framework*

14. The following are considered by the group to be the most plausible elements which could be deployed within a research assessment system for the arts and humanities, given the limitations of the available data-sets and the 'state of the art' in research impact and quality assessment techniques:

- (i) Research outputs (paragraphs 18-23).
- (ii) Spend on research infrastructure and other funding of the research environment (paragraph 24).
- (iii) Wider social, cultural and economic significance of research process (paragraph 25).
- (iv) PhD completions per research active member of staff (paragraph 26).
- (v) Peer-reviewed external research income (paragraphs 27-28).
- (vi) Esteem indicators (paragraph 29).

15. This will inevitably change over the coming years and it will be necessary for the assessment system to be dynamic and flexible enough to reflect these developments in its operation. The weightings for each element would need to be determined, bearing in mind the guiding principles outlined above. However, it is suggested that, in redressing the balance between peer-review of outputs and other elements of assessment, element (i) above should represent 50% of the assessment for any given unit. The elements (ii) to (vi) would then represent a further 50%, with the possibility that the relative weightings given to these elements (within the 50%) may vary by discipline to some degree within minimum and maximum thresholds.

16. It is recommended that the final grade given to a submission be awarded by a panel of peers based on the information provided through the peer review of outputs and the quantitative and qualitative data submitted under the elements outlined in paragraphs 18-29.

17. For each of the elements listed below, the group recognises that there are a number of caveats that need to be kept in mind when implementing them. The precise weighting for each element and the level of inputs or outputs that would achieve the highest grade would vary according to discipline-specific norms. The outcomes will vary depending on what is counted and how it is counted so that, for example, cash-volume measures of grant income may produce different results to counting the number of successful peer-reviewed awards. Wherever possible, the various elements would need to take account of the size of the submitting unit, as, for example, with total income and average income per FTE staff. The range of difficulties encountered suggests that any assessment system should have a large degree of flexibility and sensitivity to disciplinary variation.

#### *Research outputs*

18. This will remain an important element in judging research quality. Productivity should be maintained at the levels expected by the current RAE. While productivity levels should be maintained, ways of reducing the assessment burden and improving the rigour

of the process must be sought. This might be accomplished as described in the following paragraphs.

19. The current assessment burden on reviewers could be considerably reduced by a sampling of submitted outputs from individual researchers. Such peer review would be far lighter, on the whole, than the peer review used in the current system.

20. Some of the effort saved by this system should be traded against measures which would improve the rigour and credibility of such peer review. For example, more rigorous peer review may be provided by drawing upon the knowledge held within the different constituencies of peer reviewers as well as the existing systems of peer review. Existing bodies of reviewers, such as those in Research Council peer review colleges, could usefully be called upon to enhance the review of the quality of research outputs. Such bodies include specialists with a much broader range of expertise than it is possible to represent on an RAE panel; international assessors; and relevant non-academic experts. Each of these features offers a real advantage to the process of reviewing the quality of research outputs.

21. In addition to these advantages, moving the peer review of individual outputs away from the RAE panellists would also have spread the assessment burden more widely. With the burden thus spread, it would be possible to commission more than one peer review for a sampled output, hence strengthening the rigour of the system. Another advantage of using established, standing, bodies of peer reviewers is that the uniform induction and training of such peer reviewers leads to a greater consistency of reviews and outcomes.

22. The method could enhance further the credibility of the process of peer review used in research quality assessment. The current RAE assessment process, where the panel constitutes both judge and jury, would be replaced by a system whereby one or more reviewer who is not on the assessment panel is asked to review the selected output. The role of the assessment panels would be restricted to moderating the results of these reviews while balancing them with metrical data to award an overall performance rating.

23. It should be clear that such changes in the approach to peer review are not simply designed to reduce costs or mitigate the burden imposed by the assessment of outputs, but reflect a shift of emphasis in what is being assessed. The aim of moderating the current level of concentration on outputs is to enable a balanced assessment framework which captures more fully the totality of the research process.

#### *Spend on research infrastructure and other funding of the research environment*

24. In addition to reflecting a discipline's demand for, and access to, resources, this measure should take into account the total spending of the HEI on physical infrastructure and the broader research environment and the resources available to it. This could be

assessed by both quantifiable and qualitative contextual measures. Examples of such elements include (but are not restricted to):

- spend on postgraduate training
- spend on research facilities
- library spending for the submitting unit
- narrative self-assessment by the unit.

*Wider social, cultural and economic significance of research process*

25. This element refers to the significance of an output or outcome to audiences beyond the academic community. Each broad disciplinary area would develop a suite of indicators which could provide robust evidence of significance for each component of this element, and suggest the most appropriate weighting for each. The assessment of this element might require further training of reviewers. It is recognised that significance may be evidenced at all key stages of the research process.

*PhD completions per research active member of staff*

26. This element would need to be normalised according to department and institution in order to avoid unduly privileging some institutions. Consideration should also be given to student demand in emerging or unfashionable areas of research in order to avoid penalising those areas.

*Peer-reviewed external research income*

27. As mentioned above, care is required when deploying this element as a measure of research performance. It would not be satisfactory merely to aggregate the cash-volume of awards made to a submitting unit, without taking into account the demand for inputs in that discipline or the size of the unit. In addition to the actual demand for grant income, it would also be revealing to look at a discipline's capacity to access research funding. A more accurate proxy for research quality may be to count the number of successful peer-reviewed awards or activities, although this too would have to be normalised for disciplines and units and would not reflect the large number of high-quality awards that go unfunded in competition. One metric which may be used to overcome this objection is the profile of a unit's peer review scores (both unfunded and funded) over the assessment cycle, using data from Research Councils and other funders.

28. Finally, it should be borne in mind that not all peer review processes are equally rigorous nor do they all reflect the same level of esteem. Further work would need to be done to establish the weight to be applied to sources of external funding for each discipline.

*Esteem indicators*

29. It is clear that this element does not lend itself easily to quantification and may need to be subject to peer review or some other form of auditing. Furthermore, there is considerable disciplinary variation in the recognition that esteem indicators receive. This should be taken into account. The aim should be to extend and improve upon the range of indicators beyond those used in RAE 2008 where appropriate. Examples of the sub-elements that may be used include:

- invitations to give named lectures/lecture series
- invitations to contribute to, or co-author, essay collections, edited volumes or series
- organising international conferences, workshops, colloquia, exhibitions and performances, either in the UK or abroad
- membership of editorial boards.

## **Conclusion**

30. The group is not proposing a specific model for the assessment of research in the humanities. Rather, it is recommending a set of key principles and broad operational features that would inform any model that is developed by the HE funding councils post-2008. In seeking to establish a framework that can measure the totality of the research process, the group has recommended those elements that together provide the most accurate and verifiable picture of research performance in the arts and humanities. It has concluded that, at this point, metrics alone will not allow the overall performance and quality of research to be assessed at individual or departmental level.

31. Nevertheless, metrics have an important role to play in research quality assessment. Additionally, such metrics are valuable in providing the evidence to inform the judgements of reviewers, and should prove useful in building up a profile of research performance at national level for the purposes of international benchmarking. The group recognises, however, that the research landscape is evolving and that an assessment system needs to be responsive to the dynamism and innovation of the sector. The elements presented here should therefore be monitored on a regular basis to ensure they are performing their role in recognising, encouraging and rewarding world-class research.

## **Annex A**

### **Terms of reference and membership of the expert group**

#### **Background**

1. In the *Science and innovation investment framework 2004-2014*, the government reiterated its commitment to the dual support system as the basis for distributing public funds for research in higher education through the HE funding bodies and the Research Councils, and set out distinct strategic aims for each of these two funding streams. The overarching national policy is to support research of the highest quality to benefit the economy and society. The subsequent policy statement *Science and innovation investment framework: next steps* reiterated this policy and announced early action towards the greater use of metrics in allocating funds through the HE funding bodies in particular. A consultation on the means of implementing this was issued on 13 June for responses by 13 October. The consultation paper recognised that possible metrics are generally less well developed, and less straightforward in their application, in the arts and humanities and in the social sciences than for medical, scientific and engineering disciplines. It therefore called for further work to develop a more differentiated approach to recognising and rewarding research excellence in the former group of disciplines.

2. In this context HEFCE and AHRC have jointly established an expert review panel to advise on the use of metrics and other indicators in assessing quality and allocating funding for research in the arts and humanities. HEFCE is the body primarily charged with supporting funding research in the arts and humanities in England, applying approximately £200 million to that purpose in England this year. Under its Royal Charter the AHRC, which operates across the UK, is empowered to promote and support high quality research in the arts and humanities. It thus has a direct interest in the performance of arts and humanities researchers in the UK and the quality of their outputs. It is the largest funder of arts and humanities research in the UK after the funding councils and is connected with the whole community of researchers through its many panels and its peer review college with some 500 members.

3. The group has been asked to report to the chief executives of the two bodies by the middle of October 2006.

#### **Questions**

##### *General*

4. What are the distinguishing characteristics of excellence in research in the arts and humanities, and how might these be recognised and reinforced through any system of quality assessment and funding allocations to deliver the policy aim of government and the public funding bodies? What behaviours, and what types of research activity, should assessment and funding systems consequently seek to incentivise?

5. What metrics-based<sup>1</sup> approaches to assessing quality and allocating funding in the arts and humanities are possible now, or could become so in the next few years? To develop a robust and effective approach, how broad a field of potential metrics and related indicators of quality should be considered?
6. In developing a metrics- based approach:
- a. What should be the reference point for indicators of quality – national excellence, international excellence or a combination of these?
  - b. How far and in what ways can issues of dissemination and knowledge transfer and the requirements of research users be taken into account in selecting and using indicators of quality?
  - c. Is there a single metric that would meet our various needs as funders (for quality assessment, for funding or for both)?
  - d. If a single metric approach is not possible, what combination of metrics might be employed? Should these be differentially weighted? If so, how should this be achieved?
  - e. Should qualitative indicators have a role to play? If so, which such indicators might be used and how should these be interpreted in combination with the metrics?
  - f. Having established a metric or a suite of indicators, how might this best be applied in practice? Is it feasible to produce robust indicators of quality that can also be used in funding, or is some degree of distinction between quality assessment and funding allocations more workable?
  - g. Should the arrangements include a role for an oversight mechanism of the metrics-based quality criteria? If required, what form should it take?

---

<sup>1</sup> In thinking of such metrics the group is empowered to think as widely as possible, considering any possible metric except the use of the output of the European Reference Index in the Humanities project.

### Membership of the expert group

<b>Chair</b>	Professor Michael Worton, UCL (University College London)
<b>Ex officio members</b>	Professor Tony McEney, Director of Research, AHRC Paul Hubbard, Head of Research Policy, HEFCE
<b>Academic members</b>	Professor Bruce Brown, University of Brighton Professor John Caughie, University of Glasgow Professor Roger Kain, University of Exeter Professor Halvor Moxnes, University of Oslo Professor Morag Shiach, Queen Mary, University of London Professor Paul Slack, University of Oxford Professor Liz Slater, University of Liverpool

## Annex B

### Additional background information presented to the group

Item	Description
1	Consultation with AHRC Postgraduate and Research Committees, subject associations and learned societies (May-June 2006)
2	Discussion paper on most common criticisms of metrics (includes survey of press and selected academic literature and reports)
3	Summary and discussion of available metrics
4	Evidence Ltd report for OSI: PSA target metrics for UK research base (& summary of issues of arts and humanities)
5	Evidence Ltd report on International Research Activity Data
6	Evidence Ltd presentation on indicators for arts and humanities
7	Bibliometrics and citations in the arts and humanities: an overview
8	Analytical framework for AHRC evaluation strategy: 'Ecology of Knowledge'
9	<i>Setting the scene: a review of current Australian and international practice in measuring the quality and impact of publicly funded HASS research</i> , Dr Claire Donovan, Research Evaluation and Policy Project, ANU
10	<i>High correlation is not the whole story: a cautionary note on the use of simple metrics to determine QR allocations</i> , Diggle and Chetwynd, Lancaster University
11	Description of bibliometric model for performance-based funding (Norwegian Association of HEIs)