

General Comments

The purpose of the new REF, as with all previous attempts to measure research quality, is to record and assess the nature and spread of research excellence across the disciplines and act as the basis for decisions about research funding. It should not, even as an unintended consequence, alter the nature of the research that it seeks to measure. There are very real dangers, however, that the current proposals will, if implemented as they stand, result in substantial and in some cases highly undesirable changes to the nature of research and publication activity in English, and the Arts and Humanities more generally.

The RAE, it is true, has always influenced research culture to some degree, primarily through its imposition of six or seven year periods for measurement, its requirement for 'four items' and its increasing emphasis on grant income as a measure of research quality. These factors, taken alongside shifts in the policies of the funding councils, have produced significant changes in the patterns of research activity in English as well as in our sister disciplines, not all of which have been advantageous or productive. There is now distinct pressure to produce four (and only four) substantial items in any RAE period, if necessary accelerating completion to meet an RAE deadline or holding back a piece of work until after a deadline if one already has 'enough' items, rather than allowing research to emerge when it is completed to the highest possible standard. There is also pressure to be wary of long-term projects (however vital their importance) which do not attract large grants, as they may lead to 'fallow' returns in a given exercise; and contrary pressure to devise projects which do attract large grants, even if they are not always of the highest priority or utility, simply to increase grant income. All of these factors impact most obviously upon early career researchers and those in search of first posts.

But, the proposals outlined in the current consultation raise the prospect of shifting the nature of research culture in English in ways far more significant than any previously seen. Their stress on citation-indices and bibliometrics (still in their infancy and highly unreliable in the Arts and Humanities) and their movement away from the element of peer review (in the sense that it is currently understood and has operated in previous exercises) seem likely to create major adverse and unintended outcomes, favouring certain kinds of research activity and certain areas of research in English at the expense of others of equal importance, and encouraging rather than deterring a number of rather obvious forms of artificially reward-driven behaviour and game playing.

We believe that the document 'Use of Research Metrics in the Arts and Humanities' (2006) (footnoted at '53a'), is very significant. However, this document – central to an Arts and Humanities perspective – is, on occasions, revealingly at cross-purposes with itself. For example, it argues (wrongly in our view) against over-prizing research outputs at the expense of research activity ('10 viii'), but it immediately argues ('10 ix') that 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 the peer review processes as the central element in the process of Research Assessment/Evaluation if it is to retain the confidence of the community and its major

stakeholders. See 3a below, however, for our views on the vital importance of the exact nature of the peer review envisaged.

Q1a and b.

While recognising that this two-pronged question is aimed primarily at the Science disciplines (it asks whether proposals for ‘defining the broad group of *science-based disciplines*’ should be endorsed (1a) and whether there are ‘issues in relation to *specific disciplines within this framework*’ to be considered (1b)) we should state at the outset that we firmly believe that English should remain a single UoA in any future RAE.

English is a diverse, wide-ranging subject in which there is a very high volume of UK research. It formed one of the largest units of assessment in terms of the volume of published materials assessed in RAE 2008, equal in size to many groups of subject areas covered by other Panels.

Given its size and range, English is best assessed on its own, and in terms that are appropriate to the discipline. English ranges from Old Norse to Contemporary Literatures in English and Creative Writing and contains a huge range of areas of research in between – all of them ‘subjects’ in themselves that each require specialist judgement. Research in English often contains strongly multi-disciplinary elements, but its use of such elements is highly distinctive. With History, it is distinct from other Arts and Humanities disciplines in a number of its methodological approaches and in the particularly high level of importance it places on the book (the monograph, collection of essays or scholarly edition) as a principal form of research output. Merging it with other Arts and Humanities subjects for assessment purposes is likely to under-represent that distinctness and result in assessment methodologies unfit for the purpose of measuring research quality in English.

Consultation questions 2a and 2b.

Responses to these questions belong under 3 for English. However, please note that we do share the concern mentioned in ‘49’ about ‘the potential behavioural impact of a bibliometric-driven quality assessment system’.

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

The key issues that should be considered in developing light-touch peer review for the non science-based disciplines are:

1. that such light-touch peer review should be able to command the confidence of the English Studies subject community;
2. that it should not be conducted in ways that distort current modes of carrying out research, or change the nature of the discipline being assessed. For example, there should not be a move away from publishing monographs to undertaking large-scale projects simply because the latter are more likely to attract research grants, nor should subjects be grouped

- together simply because that makes available a more substantial amount of data;
3. that it should be able to deliver robust and defensible judgements about the *quality* of research.

The Meaning of ‘Peer Review’

The crucial issue here is what we mean by ‘peer review’. When CCUE and our sister subject organisations have responded to previous consultations, stressing the centrality of peer review to any Research Assessment process, we have understood ‘peer review’ to involve the assessment of research *and research culture* in English by ***a panel of English specialists***, as in the current and all previous RAEs. In these exercises, researchers in English have been assessed by their peers: that is other researchers in English, who have been able to reach nuanced judgements about the nature of an institution’s or an individual’s work (not simply on individual ‘items’) based on a close knowledge of the nature of research in English, making adjustments to the impact of metrical data, etc in the light of their knowledge of the fields or sub-fields of work concerned.

The process envisaged by the present consultation – in which expert judgement on individual outputs informs wider strategic decisions taken by broadly drawn panels of non-specialist assessors - does not in our view constitute a rigorous system of peer review.

We see no acceptable alternative to the continued use of an intensive system of close reading expert peer review as the main mode of arriving at judgements about research quality in English. If the driver for change is the burden of work on (sub-)panels, we recommend the use of more (sub-)panel members. There are strong arguments in favour of expanding the number of such members in English to ensure that research in a complex, diverse and constantly changing discipline is appropriately reviewed.

For all the limitations of the present RAE exercise, it has retained the broad confidence of the subject community because it has been based upon both specific and strategic judgements reached by a panel of experts chosen for the purpose in consultation with that community. If that confidence is to be maintained in future exercises, we strongly recommend that an expert peer-review panel of English scholars remains at the heart of the assessment process for English, and that each subject community will continue to be consulted on the membership of that review panel.

Citation Indices and Bibliometrics

The other major concern relates to the use of metrical analysis in any future exercise.

We agree strongly with the first clause in ‘53c’, namely that ‘the bibliometric data currently available for these disciplines [including Arts and Humanities] are generally not sufficiently mature to produce robust indicators’. Those who wish to introduce the use of such data need to prove that they can serve as ‘robust indicators’ of research quality. The burden of proof lies heavily on advocates of bibliometrics. Until they can be shown to work, they should not be used.

There are numerous problems with the use of citation as an index of research quality in the discipline which have thus far proven intractable, some of which we outline below.

Forms of Output

There are practical, structural problems created for a discipline in which the majority of citations do not appear in electronically visible forms (i.e. as monographs, chapters in books, scholarly editions, articles in journals not currently covered by citation engines). In time, robust methods of capturing and measuring citations in and of books and chapters *may* be developed, but they do not exist as yet.

Citation is itself a highly unreliable measure of research quality (see below), but there is an additional danger that any early move towards a greater use of citation metrics in assessment, before there is a comprehensive system of measurement available, will result in game-playing, whereby scholars will be encouraged to publish only in those forms and media which are amenable to metrical assessment. (Anecdotal evidence already exists of colleagues being told by research managers with backgrounds in Science that they should 'stop writing books' and move to short articles instead.). This would have disastrous consequences for the long-term health of the discipline.

Time-frames

In English, as in many cognate subjects, the time-lag between the production of a piece of work and the widespread recognition of its value as research is far longer than in the Sciences. Given the nature of book and journal publishing in the Arts and Humanities, it takes far longer for a work (whether an article or monograph) to move from completion to publication. This has been noted, and to some extent compensated for, in previous RAEs by the application of a slightly different time-period for assessment for Arts and Humanities subjects. But it also takes several years for a monograph or collection of essays, once published, to be reviewed in most journals (even on-line ones), let alone (given the constraints mentioned above) for it to be cited in new work. This fact has profound implications for any assessment methodology using citation-metrics that is based on the usual 6-7 year assessment periods.

The Short Term: The works cited most immediately and frequently within a conventional period of RAE/REF assessment will tend to be those which are published at the start of the exercise period rather than towards the end, and those which are either simply fashionable (and may well fade from the critical landscape very quickly) or those which are deeply controversial or simply factually inaccurate, and which the scholarly community needs to challenge or correct.

Numerous examples can be cited of works which are widely cited because they contain serious errors that it is important to correct. Such citations are measures neither of the quality nor the influence of such works. One obvious unintended outcome of a move to citation-measurement within a given REF period may therefore be that it will provide incentives for scholars who wish (or, more plausibly, will be pressed by their managers) to 'bump up their citations', to rush into print with contentious and inaccurate claims.

Any system of citation metrics must be able to distinguish between positive and negative citations. The only tried and trusted means of so doing is close-reading of all outputs by expert peer reviewers.

The HEFCE website's discussion of this issue is interesting, but does not, in our view, account for the complexities of the situation in English publishing. Under the question: "Can bibliometrics take account of 'negative' citations (highly cited research that is proven to be false)?" it offers the reply: "No, but the expert advice is that research that becomes highly cited in this way is sufficiently rare not to affect results when aggregated at the broad level of subject groups." While this may be true for some Science-based subjects, it is less clear that this is the case for the Arts and Humanities, where questions of what is 'true' and 'false' are not so easily assessed. Anecdotally at least, it seems clear that there are many forms of citation in Arts/Humanities disciplines where no great 'value' is being accorded to the work cited.

To give three instances:

1. A piece may, for example, be 'highly cited' not because it is generally regarded as 'excellent' but because it is perceived as being 'interesting' for what it may reveal 'symptomatically' about contemporary culture and society, or about the particular group or individual whose views it represents.
2. In contentious areas of debate, specific pieces may be cited not because they are generally viewed as most 'true' but because they represent articulations of the most 'extreme' positions within that debate. (The result here again could be to discourage more 'sober' contributions, insofar as more extreme, one-dimensional, responses to an issue are likely to find themselves more commonly cited.)
3. Work which is at the borderline of 'research' altogether may, in fact, be highly cited in some fields, because of the specific service it performs in the way of easy exposition. With contemporary academic publishers increasingly privileging 'introductory' texts aimed at undergraduates, these texts are often more visible and easily accessible, but they may also be used for convenience as summarising particularly complex ideas within a larger argument. Yet this does not imply any judgement of 'excellence' on the part of the citation, only one of a certain 'usefulness' or 'economic' value.

The Longer Term: Any move to embed citation-metrics into the REF must, therefore, as well as producing a far more reliable system of citation measurement and assessment than is currently available, involve a radical rethinking of the timeframe for assessment, so that citations are assessed over a far longer period. Provision would need to be made in the 2008-2012 REF exercise for work produced during the 2001-7 RAE period, or earlier exercises, as these are the works most likely to be extensively cited 'positively' during the later period.

Exclusions

The fact that only some 11% of journals in English Studies are currently covered by existing citation indices indicates the scale of the task ahead, if a system is to be developed to reflect the range of relevant citation in journals (let alone monographs, essay collections and other media) within any workable timeframe.

We also noted with some concern the statement in the Leiden study (reflected elsewhere), that journals in languages other than English were excluded from the current scoping studies. This will obviously be of major concern for disciplines within Modern Languages, but will also impact upon aspects of English language study and other areas of our discipline, where publication in languages other than English is a common.

The Nature of Citation in the Arts and Humanities

There are, however, more fundamental problems with the use of citations in English (and in many Arts and Humanities disciplines more generally) that need to be addressed before they can be used as a plausible measure of research quality. The basic principle set out under '31' in the consultation document, namely that 'the research outputs in any discipline that are most highly cited by other researchers will generally be those which have the greatest intellectual influence', is far from established as fact. It speaks of 'influence', not 'quality', and the two things are far from synonymous, especially in a discipline such as English. Even where citations can be shown to be 'positive' ones, problems arise from the nature of the discipline that render measurement problematic.

Inbuilt Biases

By its nature, the measurement of citations favours those outputs that are more widely read – factors which may have little or no relation to precise measures of their research quality (see above). This obviously favours certain kinds of output and certain fields of research. In mainstream literary criticism it favours the standard, core textual edition over the analytical article or book (for example, a Penguin edition of a Dickens novel will be cited far more (and far more quickly) than an extensively-researched monograph on medieval theatre; an article on gender in Virginia Woolf more widely than one on gender in *The Exeter Book*). It also favours the older and more established interpretation over the newer and more radical study, and the mainstream analysis over that with a more specialised focus. All of this is likely to encourage a focus on the mainstream and the canonical over the newer and more radical across the discipline.

Citation is also, of course, more widespread in those fields in which more scholars work. This again favours the mainstream and canonical over newer and marginal areas, and more modern fields (especially the 20th century) over the medieval and early modern periods and other areas in which there are currently proportionately fewer scholars. It will have a disproportionate and adverse effect upon smaller fields and sub-disciplines, many of which are of vital strategic importance (the national and regional literatures of the various parts of the UK, and Asian and Black English literatures are obvious cases in point). And it will militate against the due recognition of newer and emerging fields where, by definition, there are as yet fewer scholars working (and citing) and fewer publications in which they can be published and cited.

Citation indices also favour work published in US publications over those in the UK and Europe, reflecting the far larger academic community in the US. Were they to be used to determine funding in the UK, they would encourage UK academics to prioritise publication in US journals over UK ones, thus damaging the research infrastructure in the UK.

More fundamentally, there are certain very important areas of the discipline, both mature and emerging, that are much less amenable to citation than others. A case in point is Creative Writing, probably the most significant growth area in English Studies over the past decade, in which citation is simply not practised in any meaningful or measurable form. Many colleagues produce research-based poems, drama and prose works that will be extremely influential on others' work, but which are not 'cited' in obvious (or electronically visible) form. Hence there is no way to judge the research quality of Creative Writing outputs through citation-metrics. Given the strategic importance of Creative Writing research, especially at Postgraduate level, to English's engagement with the lifelong learning and widening-participation agenda, this seems a serious flaw in any system of measurement.

Journal Ranking

One suggested means of nuancing the purely quantitative counting of publications and citations involves the ranking of journals. The uniformly negative responses from all the Arts and Humanities communities over the recent, ill-advised ERIH journal ranking exercise should give pause before any similar system is considered for possible use in the REF process.

We note the statement reassuring us that this is not envisaged as part of the REF proposals, but journal ranking IS implicit in the Leiden study which is cited at point 26, and which underpins some of the assumptions about metrics elsewhere in the proposals. We are thus sceptical that such a system would function, even rudimentarily, without some kind of ranking being implicit within it.

We would thus point out that ranking journals in terms of status or quality does not have the same meaning in English that it does in certain areas of the Sciences. Excellent Arts and Humanities research currently appears in a wide range of journals, including those which would have no prospect of passing through the lengthy indexing process in time for the review (new publications and small-scale journals). It is clear, too, that many books will not be available for citation purposes.

One of the most obvious casualties of a citation-metric system is the promulgation of high quality research – especially perhaps interdisciplinary research - in very specialised periodicals with a limited circulation. Finding journals to take interdisciplinary work can be a real problem, and they tend to not to have large circulations. *The Sixteenth Century* and *Seventeenth Century Journals* are examples. *Women's Writing* too is a very important journal in its field, and often has articles that would be too 'specialised' for other journals. A move to measuring citation would, in effect, penalise scholars for producing

primary research in specialist fields, or for producing specialised research that provides the bedrock for more general studies.

More fundamentally, we have serious reservations about the idea of using journal editorial decisions as a factor in any transparent system of research quality assessment. HEFCE rightly notes that 'reputable peer reviewed research journals employ sound procedures for quality assuring the material they publish'. However, as a number of the editors of such journals have observed, there has to be some concern about what amounts to handing over 'quality control' to (unpaid and ultimately unaccountable) journal editors and readers. This effectively means outsourcing peer review, and placing an immense (and, in many cases, unwanted) responsibility upon those involved, insofar as their judgements would directly influence levels of individual and institutional funding for research in the future. If nothing else, there should at least be some debate about how acceptable it is for HEFCE and the government to ask of journals that they assume such responsibility.

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

We believe that the only option for the review should be that it takes the existing RAE-format, i.e. that an expert peer review panel of scholars in English should actually read the outputs and institutional statements produced by the community once more: a process which is in keeping with the very notion of the 'humanities' as an intellectual discipline.

The stated aim of lightening the touch of peer review is to reduce the burden of work for panels and HEIs. For panels, reducing the number of outputs read will clearly lighten the burden. It also, however, reduces the accuracy of finely-drawn distinctions between output quality. Moreover, if, for example, the panel read fewer than four of the items returned, who would decide which of the four should be read? Far better to increase the number of panellists and spread the burden of assessment over a wider group of specialist readers.

For the HEIs, the burden relates not so much to the number of outputs to be put forward, but to the enormous significance that is attached to the outcome of the RAE/REF. This has numerous consequences. Academics spend disproportionate amounts of time on second-guessing assessment criteria and finessing submissions; targeting supposedly highly rewarded outcomes (articles in 'the right journals', 'super-books', big grant capture); hours are spent scrutinizing and grading the proposed outputs of colleagues; much effort goes into determining which staff should be entered, with complex procedures to protect the right of appeal for those denied that privilege; departmental descriptions are drafted and re-drafted with ever closer attention to minute detail; intricate electronic databases are designed and maintained.

Given the stakes involved, it does not seem likely that the proposals presented in this consultation document would reduce the pressure on HEIs to devote such time and effort

to nuancing their applications, and game-playing in order to secure the best possible outcomes.

There are practical as well as political reasons for this claim. Any upgrading of the use of quantitative data must require narrative explanation about the meanings and contexts of such data. This will create more work for those preparing submissions. For example, if citations are to be used, material explaining the significance of these citations will have to be included in the equivalent of RA5 or in the 'other' section of RA2 items. The result will be a shift away from review of RA2 outputs to greater reliance on RA5-type narrative about the meanings of quantitative data. This shift does not seem to us desirable.

The current system of looking at figures to do with grants, postgraduate numbers and completion rates, in the context of an assessment weighed heavily in favour of Outputs, but taking into account 'Environment', seems as nuanced a way of using figures as any being proposed.

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

We believe that any attempt to 'capture...research excellence' other than by the careful and painstaking act of reading is anything but robust in a text-based discipline; hence the need for the retention of the current system of subject-specialist review (sub-)panels.

In general terms, quantification rewards size. English departments come in different shapes and sizes. The small-to-medium but excellent department will suffer under REF unless metrics such as research student numbers and research income are related to the size of department. (This wasn't the case, for example, with the new BGP postgraduate funding scheme, with worrying implications for smaller departments). Like most recent changes in HE funding, these proposals are likely to have the effect of rewarding the large and historically well-funded departments, regardless of their current status and qualities.

We do note, however, the absence in these proposals of references to most of those factors currently covered under 'Esteem indicators' in RAE 2008. Such things as journal editorships, external examining roles, membership of learned societies, etc., are amenable to quantitative analysis, and, while they clearly favour more experienced researchers over early-career academics, etc, they do have some relationship to the relative impact and influence of researchers working in English. Such data might usefully be added to any metrical analysis employed by the REF, albeit it should always be subject to the scrutiny and judgement of an expert peer-review panel.

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

The proposals downplay the element of expert peer reviewers to that of external examiners of specific outputs. If peer review is to carry the confidence of the community, and produce meaningful judgements on the quality of both the research produced and the culture(s) which sustain it, panels of subject-specialists nominated from within the subject communities through transparent systems of nomination and appointment must be integral to the whole REF process from the discussion of its first principles to its operation in practice.

We would wish to retain, even increase the size of, the sub-panel in English; we do not feel it appropriate for English to be merged with other Arts and Humanities disciplines, nor for expert judgement to be relegated to a subsidiary element in the review process.

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

See also our comments in response to Question 3b in this context. Given the importance of the exercise to HEIs, it is unlikely that ANY system of research review will prevent them from instituting lengthy, burdensome and time-consuming internal processes aimed at maximising their chances of success. This will inevitably divert time and energy that might otherwise be spent on research and teaching development towards preparing for the next assessment exercise.

Looking at the question from the assessors' perspective, it is true that asking a panel to sit down and read the outputs is 'burdensome', but this is a necessary process if the quality of research is accurately to be measured. If we are to have a proper review process that assures the stake-holders of the 'quality' of research in English, then it has to be burdensome; otherwise it is simply slipshod and inaccurate.

There are serious implications for the burden on the sector of any move towards a light-touch peer review system with a significantly enhanced emphasis on metrics. We are quite sure that the institutional workload will be substantially increased, both in relation to the labour of data-retention and to that involved in explaining the significance of data. This will easily outweigh in person-hours the slight saving in time made by panels possibly reading fewer outputs (which for academic reasons we do not favour).

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

The impact on equal opportunities of the proposals made in the document could well be

disadvantageous, in that this new system seems intent on measuring every aspect of research activity (while reducing the value placed on outputs), thus leading to an increase of pressure and stress in people's daily working lives, an increase which may well be felt most keenly by (for example) colleagues on temporary or part-time contracts.

More generally, it seems important to us to digest the EO implications of the current RAE system before we move to the new dispensation. Has HEFCE done an EO audit on research funding allocations? It would seem imperative to discover whether, for example, there is evidence that female academics tend to earn less research income than their male peers, or that certain research areas associated with particular ethnic groups tend to be less successful in research funding competitions. Only then can this question be addressed with any confidence.

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

It is difficult to avoid the conclusion that, for English Studies, a change to light-touch peer review is fraught with dangers that may imperil academic and public confidence in the value and use of the research exercise. The current system (or one with more panel members) may have its imperfections, but it is more robust than the alternatives which are beginning to shape themselves in the consultation document. There seems to us no substitute for an exercise that works in its own evaluative terms rather than relying on terms derived from different contexts and activities.