On structuring subjective judgements: originality, significance and rigour in RAE 2008

Ron Johnston, School of Geographical Sciences, University of Bristol, Bristol BS8 1SS

Contact: R.Johnston@bristol.ac.uk
On structuring subjective judgements: originality, significance and rigour in RAE 2008

Abstract

The 2008 United Kingdom Research Assessment Exercise will involve the evaluation of thousands of individual research outputs. The Funding Councils set three criteria for those evaluations – originality, rigour and significance – and required each output to be placed into a fivefold categorisation of excellence, using absolute rather than relative judgements. The various panels and sub-panels convened to undertake those evaluations were asked to provide their own interpretations of both the criteria and the categories. This paper analyses those, and finds that they are no less subjective – and hence vague – than the generic descriptions provided by the central organisation.

Academics are used to grading a wide range of materials, from student essays and examinations through research grant applications, potential journal articles and book manuscripts to the materials presented by candidates for appointments and/or promotions. Some of their judgements are binary – accept/reject; some involve relative judgements – placing candidates/materials in order of merit; and some involve absolute judgements – giving a grade or mark against agreed/pre-determined standards and/or criteria. Most are subjective – the exceptions involve technical material for which there is either a right or a wrong answer. Even where so-called absolute standards are provided, nevertheless there are often no clear-cut criteria which allow an uncontestable decision as to which grade is deserved. The decision-makers very often operate with regard to identifying excellence as did U.S. Supreme Court Justice Potter Stewart who could not define pornography but knew it when he saw it.

One arena of academic judgement with far-reaching consequences which has attracted massive attention beyond as well as within the United Kingdom academic community over the last two decades, and which is dominated by the concept of excellence (on which see Allan, 2006), is the Research Assessment Exercise (RAE). Originally designed by members of the then (academic-dominated) University Grants Committee to respond to government ministers’ questions regarding the returns on investment in university research, it was (in the current context) initially a relative light-touch exercise which graded departments (or Units of Assessment: UoAs) in 1986 on a four-point scale. Differential funding for research (known as QR) in the universities followed these grades – with a relative shallow slope connecting the lowest to the highest graded UoAs. Across the next four exercises (1989, 1992, 1996, 2001) the grading process became more intensive, the scale increasingly long, the amount of material presented more extensive, and the commitment of resources much greater (all part of what Howie, 2006, calls ‘drowning by numbers’ for UK universities). And, most importantly, the differential allocation of QR moneys became much greater;
very large sums of money were at stake and much (including funding for posts and other infrastructural resources) depended on the grade achieved.\(^4\)

One consistent element across those five RAEs was that the grading related to UoAs – in most cases separate university departments – and although the panels making the judgements used materials from and about individuals the outcome was always a grading of a collective, not of an individual. Following a review chaired by Sir Gareth Roberts (2003), however, the 2008 RAE will differ from that procedure in one important respect. Although the results will still refer to a collective – the UoA – it will not be a single summative judgement but rather a profile reflecting the relevant panel’s views of the materials presented to them, in three main categories – research outputs, research environment and esteem indicators. Central to those profiles will be the panels’ ratings on a five-point scale of each of up to four separate items submitted by each researcher entered in her/his UoA. Individuals’ work is thus explicitly to be graded. Furthermore, the gradings will be the dominant determinants of the final profile. Each panel could determine the percentage weight it would give to the three categories evaluated: the recommended minimum for the outputs was 50 per cent, but most of the 15 panels decided on either 70 or 75, with Panel G deciding on 50 and Panel N on 80.\(^5\) The assessed quality of the research outputs is clearly the determining factor on the overall profile, with major implications for research funding of universities over the following years.

The RAE is a further example of academics being involved in making judgements of the quality of submitted work. Those judgements are supposedly made against absolute standards, but they will necessarily be subjective – as with all such exercises – and have to be made against stated (if not defined) criteria set out by the Funding Councils which organise the RAEs.\(^6\) Their subjectivity is recognised in that each main panel and sub-panel involved in the RAE2008 judgement process has been required to set out its own definitions in the light of the generic criteria. This essay evaluates those definitions to see what light they throw on academics’ criteria for what determining what constitutes world-class research.

Setting the stage

The RAE evaluations are undertaken for the four UK University Funding Councils by selected panels and sub-panels whose memberships are dominated by practising academics. There are 15 main panels (Table 1), each with a number of sub-panels covering individual UoAs; the role of the main panels is to ensure consistency of decision-making across a group of related disciplines represented by the sub-panels – of which there are 67 in total.

The role of those panels and sub-panels with regard to the assessment of each submitted piece of output (which need not be a publication, though most undoubtedly will be: other categories of output acceptable include images, devices, patents and software) is set out in an introductory document (RAE 2008, 2005, #26):

\[
\text{This is an expert review exercise. Sub-panel members will exercise their knowledge, judgement and expertise to reach a collective view on the quality profile of research described in each submission, that is the proportion of work in each submission that is judged to reach each of five quality levels from 4*}
\]
to unclassified (see Annex A [this is published here as Table 2]). The
definition of each level relies on a conception of quality (world-leading) which
is the absolute standard of quality in each UOA. Each submission will be
assessed against absolute standards and will not be ranked against other
submissions.

The grade descriptions reproduced here in Table 2 are further elaborated in a series of
paragraphs:

1. Sub-panels will use their professional judgement to form a view about
the quality profile of the research activity described in each submission, taking
into account all the evidence presented. Their recommendations will be
endorsed by the main panel in consultation with the sub-panel.
2. ‘World-leading’ quality denotes an absolute standard of quality in each
unit of assessment.
3. ‘World leading’, ‘internationally’ and ‘nationally’ in this context refer to
quality standards. They do not refer to the nature or geographical scope of
particular subjects, nor to the locus of research nor its place of dissemination,
for example, in the case of ‘nationally’, to work that is disseminated in the
United Kingdom of Great Britain and Northern Ireland.

It was, however, left to the panels and sub-panels – if they wished – to refine the key
terms in those definitions: world-leading; internationally excellent; recognised
internationally; and recognised nationally – as well as the criteria of originality,
significance, and rigour. The refined definitions were published by the panels and
sub-panels and form the criteria against which their judgements will be made.

**Have we any models to deploy?**

While the RAE may be relatively novel, nevertheless there has been a great deal of
academic study of disciplinary histories, which should provide some pointers as to
what work might be considered world-leading, internationally excellent etc. Of those,
undoubtedly the most influential – and the most cited (itself a measure of world-
leadingness?!) – is Thomas Kuhn’s (1962, 1970) *The structure of scientific
revolutions*.

Kuhn divided scientific activity into two main types – normal science which, as the
name suggests, predominates, and extraordinary, or revolutionary, science. *Normal
science* involves the adoption and application of a paradigm, his descriptive term for a
community of workers who operate within an agreed philosophy, concur on the
theoretical focus of their work, and use accepted methodological procedures. Once
socialised into a paradigm, members of that community are aware of what it accepts
as (albeit provisional) knowledge (the ‘known, knowns’), what problems are
unresolved (the ‘known unknowns’), and what methods are available for tackling
those problems. The paradigm is thus an ‘accepted problem-solution’ framework
(Barnes, 1982, xiv) and provides the context for normal science. Researchers operate
within this framework, tackling problems sequentially and adding to the store of
knowledge as they resolve them. Such resolutions may not be straightforward and
may call for substantial perseverance, skill and ingenuity but they only advance
knowledge within an agreed framework – by, as Newton’s much-quoted phrase indicates, ‘standing on the shoulders of giants’.  

Normal scientific practice thus involves the continued application and extension of accepted methods within an agreed theoretical context: researchers move the agreed agenda forward – some more successfully than others, some making more substantial contributions than others. (Barnes – 1982, 49 – calls what normal scientists do ‘filling in the gaps’: making ‘the unknown into an instance of the known, into another routine case’.) By contrast, extraordinary or revolutionary science involves breaking the bounds, shifting the agenda. In Kuhn’s terms, some normal science throws up anomalies (or rogue solutions), which do not readily fit into the agreed framework. These remain as puzzles in themselves (‘unknown unknowns’), instances of where the paradigm has failed. They may be (largely) ignored, as most studies within the paradigm continue to succeed. But as a volume of anomalies builds up, so concern grows about the paradigm’s viability – either to one or a few workers only or to the community more generally. This logjam is only broken when a new paradigm is created, which can incorporate not only the existing knowledge but also (some at least of) the anomalies. This new paradigm then has to be ‘sold’ to the community; the sceptical have to be convinced that there is a ‘better way’ – and this usually involves a ‘leap of (individual and/or collective) faith’ on their part, because of the incommensurability of the old and the new paradigms (the lack of common standards against which to judge an approach’s validity).

Here, it seems, is a model that could be deployed to evaluate the submitted outputs in an RAE. Originality and significance clearly apply to works that stimulate a paradigm shift – and such a shift is only likely to be successful if it is rigorously sustained. There are, however, several problems with that argument.

The first (1962) edition of Kuhn’s book took no account of what he later adopted for the second (1970) edition, following a critique by Masterman (1970), who identified no less than 21 different usages of the term paradigm in the original and suggested that these could be classified into three main types:

1. The metaparadigm, which can be equated with a ‘world view’ or general organizing principle for a scientific activity;
2. The sociological paradigm, which is a community’s accepted scientific achievements and working habits; and
3. The artefact or construct paradigm, which is a classic tool providing the model, or exemplar, for future work.

The second provides the structure within which scientists work, and the third their problem-solving toolbox; the first comprises their overall views of the nature of science and its objects of study (its epistemology and ontology). Kuhn accepted this representation of his work, and argued that most of the shifts he was concerned with involved either sociological or construct paradigms. This clearly implies a hierarchy of changes. The first involves changes within the toolbox, introducing new ways of working – new exemplars of ‘best practice’ – which occur within a community’s shared view of what they know and have yet to discover. The second involves changes in the community’s accepted knowledge base and ways of working – its disciplinary matrix. And the third – or highest – level involves changes in the metaparadigm, in the entire philosophy of science.
Using this model as the basis for an RAE evaluation, therefore, would involve establishing whether a piece of work is normal or revolutionary science and, if the latter, at which scale. Is a change of exemplar as significant as one of disciplinary matrix, let alone one of world view? Furthermore, changes at the level of the exemplar are presumably much more common than at the disciplinary matrix: we are more likely to develop a new tool to enable us to do better that which we are already doing than to introduce (and get accepted) a whole new way of approaching our subject. How many metaparadigmatic changes are likely to occur within the 6-7 years of work covered by an RAE? Or even changes in the sociological paradigm? So are the really significant and original pieces of work likely to be those that introduce a new exemplar, a new way of working within agreed limits, and the new findings that follow – unless, that is, there are to be extremely few outputs graded 4*?

The second major critique of Kuhn’s argument that is very relevant to the RAE concerns its applicability across disciplines. It presents a very linear model of scientific ‘progress’, one that may apply to the natural sciences and related disciplines but less so to the humanities and social sciences (or at least to large parts of those disciplines). Kuhn himself recognised this, in essays published in 1977 and 1991 (the latter is reprinted in Conant and Haugeland, 2000), which responded to a paper by Taylor (1985) that drew a clear distinction between studies of natural phenomena and human actions: the former can readily apply methods of scientific experimentation leading to explanation, whereas the latter can only hope to achieve hermeneutic interpretation (or understanding) of a changing social world in which all knowledge is therefore contingent. While partly agreeing, Kuhn nevertheless argued that all concepts – of the natural and human worlds – are cultural, shared by communities as influences on how they study their chosen objects. (As he put it, ‘the heavens are not culture-independent’.) And those shared meanings can be subjected to paradigmatic shocks – whether they refer to the natural or the human world; indeed a ‘natural scientific community’ is itself a culture whose traits are transmitted to new members in the same way as those of studies of human cultures by anthropologists.

Nevertheless, there is a major difference: natural sciences are not hermeneutic enterprises in the way many social science and humanities disciplines are: the former seek better explanations of the phenomena they study through accepted ‘scientific (usually experimental) methods’ whereas the goal of the former is newer and deeper interpretations.

Given that difference, how do you identify a world-leading, original, rigorous and significantly deeper interpretation? What is progress within hermeneutic studies? And just as there are different cultures operating within the world that we study (at different scales of difference) there are also different cultures, or knowledge communities (Pinch, 1998), operating within an academic discipline – and how do you evaluate contributions to one vis-à-vis those in another? Whereas most natural scientists can identify progress, and perhaps agree substantially on its importance, within a normal scientific framework, the same cannot be said of most academics in the social sciences and the humanities. Thus, according to Fuller (2000, 228):

Kuhn concluded that the history of the social sciences has not witnessed a clear succession of paradigms because social scientists have been unable to agree on research exemplars to underwrite the activity of normal science … the peculiar character of the social sciences can be traced historically to the
social scientists having been guided by larger, conceptually unwieldy social problems that cannot be reduced to well-defined puzzles. As a result, those disciplines (plus the humanities) are likely to be characterized by much more fluid structures, comprising sub-communities with fuzzy boundaries and changing memberships within which the significance of individual pieces of work may be difficult to assess (especially for ‘outsiders’).

Following that argument, Geertz (1983, 157) suggests that these ‘effective academic communities are not that much larger than most peasant villages and just about as ingrown’, working on particular subject matters or sets of problems, with particular methodologies. This may not be anarchy – as strictly defined – but it may not be far from it. And as such, how can there be common criteria for evaluation – let alone agreement on what is the most significant (and original, and rigorous) work.

Revolutions are thus rare within academic study – at whatever scale – and in some disciplines, notably in the humanities and social sciences, where there are revolutions they do not conform to the Kuhnian model as applied in the social sciences: rather than say here is a better way of explaining an agreed set of phenomena they are more likely to involve claims that the discipline should study different phenomena (and, almost necessarily, in different ways). Thus whereas some disciplines may comprise a single community with a common world view and disciplinary matrix plus a number of exemplars being used to lead research on various topics, other disciplines may contain two or more separate communities with competing world views and disciplinary matrices, let alone exemplars – and competition among them may, in effect, lead to some communities questioning the validity of others’ work (as explored in Johnston, 2006a). This leads to problems in reaching a collective determination of what is rigorous and significant even if there is agreement on its originality.

For some disciplines, therefore, a summary of the applicability of Kuhn’s ideas might be ‘paradigms yes, revolutions no’ (on which see Johnston and Sidaway, 2004). They comprise a number of distinct, if sometimes ill-defined, communities which combine, politically if not intellectually, as a coherent body of scholars committed to a particular focus – most of them on subject matter. Those communities may share a great deal in terms of world view and disciplinary matrix, but each operates with its own exemplars that are relevant to the subject matter studied. Occasionally the exemplars will change – either because of a major discovery (Chadwick’s, 1932, discovery that atoms contained neutrons as well as electrons and protons, for example) or because better technical procedures are available (as with the introduction of instrumental variables to deal with endogeneity issues in econometrics; Morgan, 1990). And even less occasionally, there will be even bigger revolutions – the Copernican/Galilean revolution in astronomy, for example, and the Einsteinian in physics.

But in a number of (mainly non-natural science) disciplines the communities share much less – there is more than one disciplinary matrix, if not world view, not only sharing but usually competing for intra-disciplinary space. Those communities not only have very different exemplars but also different agenda and goals – which suggests not only very different conceptions of originality, rigour and significance but also, leading from those, different judgements of what is world-leading, internationally excellent etc. Several paradigms co-exist – each probably associated
with a number of small sub-communities (Geertz’s ‘peasant villages’) – and may well
be in conflict. One community may however dominate a sub-panel’s evaluations –
which is what Lee (2007) explicitly argues for economics at previous iterations, with
‘heterodox economics’ degraded and marginalised because its adherents are unable to
publish in what are considered the discipline’s ‘top journals’, and Smolin (2006)
implicitly argues for physics with string theory having ‘taken over’ the discipline in
recent years. If this is not to be the case, then RAE panels and sub-panels will have to
accommodate these, often very considerable, differences within a discipline by
ensuring that some communities (and thus, potentially, university departments) are
not privileged over others in their interpretations of the criteria. Indeed, the issue
precedes these considerations, since the constitution of sub-panels needs to be
consistent with the recognition of separate communities within the discipline(s)
contained within many of the UoAs: if the sub-panel membership is considered
unrepresentative (implying greater internal homogeneity than is really the case), so
that certain types of work cannot be ‘properly assessed’ by those undertaking the task,
then certain sub-communities could be disadvantaged, with consequences for the
funding of the universities where they are most numerous.

There is no model of how academic work is undertaken and ‘progresses’ that can
readily be applied across all disciplines (or even within disciplines), therefore.
Academics operate within paradigms, communities of scholars with shared
understandings of knowledge, approaches and methods from which spring
appreciations of the relative quality of work in their (sub-)fields – in terms of its
originality, rigour and significance – and there are few revolutions, even at the
exemplar scale, within any short period. The task of the panels and sub-panels in
preparing for their assessments was to make those understandings and appreciations
specific – reflecting their, discipline-specific in most cases, separate discourses – so
that their procedures for classification of the submitted outputs would be generally
accepted. They were asked to identify the main paradigms within their respective
UoAs and specify how their subjective decision-making would be both coherent and
consistent. How did they do it?

So what is original, significant and rigorous? What is world-leading?

The production of the panel and sub-panel definitions (some panels produced a single
set embracing all of their sub-panels) involved several hundred senior academics –
who also consulted widely. The result should be a series of clear, discipline-specific,
statements of what comprises academic work of different quality. But does it? The
following analysis suggests that all that most of the panels and sub-panels have done
is to rewrite one set of relatively vague descriptors into another: the words may be
different but the imprecision is just the same – like pornography, academic excellence
can be seen but not defined.

Original, significant and rigorous

Throughout the documentation for RAE2008 great emphasis is placed on the three
criteria against which all outputs and other indicators should be assessed. This trilogy
of originality, significance and rigour is introduced in a table setting out the
definitions of the five chosen quality levels, published in the initial guidance to panels
(RAE2008, 2005, 9). Earlier in that document a more general statement indicates that:
Examples of indicators used to judge research output, for example, might include originality, imaginative range or significance, as demonstrated by the extent to which knowledge or understanding in the field has been increased or practice has been or is likely to be improved. A number of the panels and sub-panels (especially in the humanities and social sciences) have taken up the criterion of imaginative range – though none defines it – but all necessarily focus on originality, significance and rigour.  

But what do they mean? For most of the panels they are taken as self-explanatory and there is no explication at all of how they will be interpreted. This includes almost all of those covering medicine, the natural sciences, and engineering. There are a few exceptions where the terms are slightly extended. UoA 8 (‘Primary care and other community based clinical subjects’) includes a statement (#13) that ‘Outputs presenting a new hypothesis will be considered equally to primary research, and assessed in terms of their originality and power to stimulate new work’ and the four UoAs under Panel C (see Table 1) all include the following (#21 in UoA 10):

In considering outputs the sub-panel will use its academic and professional judgement to assess their quality using such criteria as:

a. Novelty and originality of the underlying idea, hypothesis and/or method;

b. Scientific rigour with regard to design, method and analysis;

c. The logical coherence of argument;

d. Significance of work to advancing knowledge, skills, understanding and

e. Scholarship in theory, practice, education, management and/or policy;

f. Engagement with service users, the public, industry and/or policy;

g. Applicability and significance to service and research users.

These, in effect, do little more than extend and amplify the list of criteria.

For the scientists and technologists involved in the RAE, therefore, it seems that the concepts of originality, significance and rigour are very largely unproblematic. Like the Justice Potter Stewart they know it when they see it and assume that their colleagues share their appreciation: there is no need to define what is self-evident.

This is not the case with many of the social science and humanities disciplines, however. Two Panels (L and O) and a substantial number of sub-panels found it necessary to set out their understanding of the three concepts, some of them in considerable detail: their texts are reproduced in Appendix I. Some provide little more than brief amplification: that adopted by Panel L, for example, merely adds a criterion (imaginative range – which is undefined) and expands on significance, whereas the three definitions produced by UoA 39 (Politics and international studies) do little more than recognise the disciplines’ plurality of approaches (which may be incommensurable). Others are more detailed and prescriptive – as with those for the four, economics-based, UoAs in Panel I which implicitly constrain the criteria to a particular approach (on which see Lee, 2007) and the long definitions adopted by UoAs 44, 45, and 46.

For each of the criteria, these explications carry very similar messages. Originality is associated with innovation – addressing new questions, producing new evidence and insights and developing new syntheses of existing work. It is virtually synonymous with novelty, though the importance of any piece of work has to be judged by the
other criteria as well. Rigour is linked to robustness of argument and method, and includes methodological advances in some cases.\textsuperscript{14} And the usual extended definition of \textit{significance} relates to agenda-setting – for both intellectual and policy-oriented work. (From its origins in the mid-1980s, part of the rationale for the RAE has been to demonstrate the applied value of university research.)

Whether any of these definitions adds a great deal to the general appreciation of the three criteria – and thereby aids individual researchers in deciding which work to submit and sub-panels how to judge it – is open to doubt. Clearly most of the academics involved in the sub-panels associated with humanities and social science disciplines considered it necessary to spell out how they would deploy the criteria – perhaps for no other reason that to convince members of their respective disciplines that they would not be overly prescriptive in what they considered of high quality by recognising the pluralism – and sometimes contested nature – that characterises many of those disciplines. If it is robustly argued, innovative and agenda-setting for at least some practitioners, then it qualifies. But scientists and technologists saw no need to undertake such (defensive?) exercises: for them the criteria were unproblematic.

\textit{World-leading, excellent}….

The five quality levels … apply to all UOAs. Some panel criteria statements include a descriptive account of the quality level definitions, to inform their subject communities on how they will apply each level in judging quality. These descriptive accounts should be read alongside, but do not replace, the standard definitions. (RAE2008, 2006, #19).

In their preparatory work, each of the panels and sub-panels was asked to consider how they would interpret the generic criteria for evaluating research outputs (Table 2) and produce a document outlining the procedures they intended to adopt. Most decided to at least put a particular gloss on those generic criteria – the main exceptions were Panels A and B (which cover the main medical sciences, excluding dentistry, nursing and pharmacy). Panel A merely stated that it would look for ‘evidence of scientific rigour and excellence, originality, novelty, potential applicability to human health, applicability and significance to health service and research users, significant addition to knowledge and to the conceptual framework of the field, and contribution to the development of the researchers of the future’: Panel B said even less, indicating that a sample of outputs would be assessed by sub-panel members to ensure consistency in application of the quality levels. A few others just repeated the generic descriptors (as with Panels G and H, although in the case of the latter each of its sub-panels then added its own gloss – in part to reflect professional, practice-based considerations as for UoAs 30 and 31).

The remaining panels and sub-panels produced a range of different interpretations and amplifications of the generic criteria, although there were some similarities – members of the RAE secretariat almost certainly suggested particular wordings that had been deployed elsewhere. Some provided very long descriptions; others were relatively short. They have not been reproduced here and the following discussion is based on a summary classification into three main groups.
The first of these groups, comprising Panels C, D, F, K, and N, emphasises the degree of significance or impact. Panel C, for example, uses the terms ‘an outstanding contribution’, ‘a highly significant contribution’, ‘a significant contribution’ and ‘a contribution’ in its descriptive sentences for grades 4*, 3*, 2* and 1* respectively. Panel K similarly employs ‘highly significant’, ‘significant’, ‘recognised contribution’ and ‘limited contribution’: for ‘highly significant’ it further states that the research will have generated ‘new methods, new practices, new theoretical frameworks, new understandings’, while 3* and 2* contributions are to ‘match the standards of internationally peer reviewed research’ – with 3* contributions in addition being of ‘high quality’. (Recall that within Panel K, UoAs 44 and 45 also provided lengthy descriptions of originality, rigour and significance: Appendix I.) Panel D deploys similar terminology as well as referring to ‘setting the research agenda’ (4*), adding ‘important knowledge, ideas or techniques … likely to have a lasting impact’ (3*), and having ‘incremental and short term influence’ (2*): work graded 1* ‘is likely to have little influence’. Finally, UoA 20 within Panel F describes the four grades of work as having ‘a significant impact’, ‘a clear impact’, ‘an impact’ and ‘unlikely to have more than a minor impact’ on the development of the field respectively, and the other sub-panels repeat that terminology.

A second group of panels – I, J, M, N and O – stresses one aspect of a work’s significance – the degree to which it is a point of reference for other workers. Panel I, for example, describes 4* outputs as ‘comparable to the best work in the field or sub-field … It has become, or is likely to become, a primary point of reference in its field or sub-field’. Moving down the scale, 3* outputs will be, or likely to be, ‘a major point of reference’ whereas 2* and 1* items will make ‘a contribution’ or a ‘limited contribution’ respectively. Panel J uses the same terminology for 4* and 3*; 2* are ‘a reference point’ and 1* ‘a contribution’. In Panel M the four descriptors are set out in each sub-panel’s document, using the same terminology: 4* refers to an output that …is, or ought to be, an essential point of reference in its field or sub-field, and makes a contribution of which every serious researcher in the field ought to be aware whereas 3* outputs make a contribution of which every serious researcher ought to be aware; 2* outputs are ‘substantial contribution[s] … likely to inform subsequent work’. Panel N uses very similar terms except for opining that 2* work ‘merits attention in the field’ and 1* ‘merits some attention’. Within that panel, the sub-panel for UoA62 (History) expands on those descriptors – notably for 4*:

Originality can be in the form of reshaping interpretations or approaches or opening up new sources, new data or material. Significance will be judged on the basis of, for example, depth and likely lasting scholarly value. Rigour will be judged on, for example, accuracy, clarity and standards of scholarship. Work graded 4* will be outstanding in respect of virtually all these qualities. Finally, for Panel O the respective adjectives to ‘point of reference’ for the four grades are ‘essential’, ‘major’, ‘important’ and ‘useful’.

The third class, in some ways a subset of the previous one, is the paradigm shifters. Kuhn’s concept imbues Panel E’s lengthy grade descriptors. Thus 4* work in the core natural sciences of physics, chemistry and earth systems is ‘agenda setting .. leading or at the forefront of the research area … [demonstrates] great novelty … [a] major influence on a research theme or field … developing new paradigms or new concepts for research … [and stimulating] major changes in policy or practice with regard to
applied research’. By comparison 3* work ‘contributes important ideas and techniques which are likely to have lasting impact, but are not developing new paradigms’, and 2* outputs use ‘established techniques or approaches, which conform with existing ideas or paradigms’. This is by far the most developed use of Kuhn’s terminology. Panel L’s descriptors have much in common with those in the point of reference group, but – perhaps surprisingly for area studies disciplines normally associated with the humanities and social sciences – its sentence for 4* begins with ‘Research of the highest standards which is paradigm-shifting or which in the panel’s informed judgement is likely to become so…’. Finally, within Panel H, whose general statements emphasise agenda-setting, 1* work is described as ‘contributing to existing paradigms and agendas’: UoA 32 (geography) extends this use of Kuhn’s concepts, describing 2* work as ‘offering an incremental advance within existing paradigms, traditions of inquiry, or domains of policy and practice’ and 1* as ‘largely framed by existing paradigms…’.

Kuhn’s model of disciplinary change is only explicitly deployed in a small number of cases, therefore. Even in those, the scale at which the concept of paradigm is used is far from explicit, although it is hard to expect shifts other than that of the exemplar – Panel E refers to new paradigms (note the plural!) and ‘fundamental new concepts’ and Panel H to ‘existing paradigms and agendas’. It seems likely that although the term is used, the details of its definitions have not been carefully addressed. The deployment of paradigm for 2* and 1* descriptors in Panel H clearly implies working within the frameworks of existing research communities: ‘paradigm-shifting’ and ‘new paradigms’ for N and E respectively imply the establishment of new research communities. How likely is it that one piece of output will do that?

What comes through more clearly from not only those using the paradigm concept but also the grade descriptors drawn up by other panels and sub-panels is that the academics involved have been unable to produce descriptions of the four grades that improve substantially on the generic set provided at the outset of the exercise. Academics frequently make subjective judgements of the academic quality of work submitted to them – by their academic peers (through the journal refereeing process) as well as by their students – even though they have little or no training to do that: as Middlehurst (1993) has argued, there are seven ‘cults’ of academic life – of the gifted amateur, heredity, deficiency, adequacy, implicit, selection, and individual – which are used to deny the need for training in a range of tasks, and evaluation is one of them. They bring that experience to the RAE: they are able to judge what is likely to become a primary point of reference in their field, what will have lasting impact etc – and they may well be right (how could they be put to the test?).

One important aspect of many of the definitions is the implication that they carry for the development of ‘objective indicators’, which the Funding Councils, under significant pressure from the government, are committed to use for future RAEs, replacing the time-consuming (and therefore expensive) subjective judgements by metric-based systems (on which see Johnston, 2006b). Thus, for example, the statement by Panel K that 3* and 2* contributions are to ‘match the standards of internationally peer reviewed research’ could well mean that publication in a defined set of journals is considered a mark of that level of excellence (which is what Lee, 2007, and others have claimed occurs in economics evaluations), as well as successful grant-winning from recognised bodies such as the UK Research Councils. Several of
the panels and sub-panels explicitly claim that they do not have lists of the ‘more
prestigious’ journals (see note 10) and claim that the quality of the work will be
assessed by its internal merits, irrespective of where it appears. Nevertheless, the
implication is there: research of a certain quality meets the standards of certain outlets
and funders – even though those too could not be expressed in other than subjective
terms like those discussed here.

A further indicator of the potential for a shift towards a metrics-based RAE is the
number of references to outputs that have made or are likely to make (highly
significant, significant, recognised etc.) contributions to their field (another undefined
term), as well as to (primary, essential etc.) points of reference, to every serious
researcher in a field being aware of an item, and it informing subsequent work. Such
statements clearly imply that the outputs are or will be heavily cited within their field,
because many other researchers will build on them – developing their ideas, testing
their hypotheses, deploying their methods etc. Citation counts are already widely used
in academic evaluations, with the frequency of citation being taken as a major
indicator of the impact of a piece of work – although the volume of citations is rarely
linked to the number of practitioners in the relevant field/sub-field (as illustrated by
Bodman’s, 1991, 1992, work). They are, of course, retrospective indicators and it may
be difficult to use them for recently-published items whose impact is not immediate –
which may well be the case in some disciplines more than others – but the implicit
linking of the grade descriptors to existing and potential metrics gives a clear clue to
how many academics think about the impact of a piece of work, and possibly an
indicator of what some of the RAE2008 sub-panels’ evaluators may do when
classifying the various outputs. (On the nature and value of citation counts, see
Yeung, 2002.)

Conclusions

Academic life – as with many other aspects of society – is replete with individuals
making judgements about either other individuals or examples of their work,
judgements that can have major consequences regarding careers and even life-
chances. The UK Research Assessment Exercises are an excellent exemplar of an
institutionalised judgemental process within academia with major impacts not only for
individual careers but also the financial and intellectual health of groups of
researchers (particularly academic departments) and even entire universities and other
institutions. Because of the perceived importance of those judgements – and the
possibility of them being challenged legally – those managing the latest RAE have
sought to formalise the criteria that will be deployed and the outcomes that will result,
in a new system that involves the explicit ranking of every item of research output
submitted to the exercise.

To a large extent, the formalisation has failed. The overall schema involved the
identification of three major criteria – originality, rigour, and significance – that are to
be used to place research outputs in one of five grades according to their perceived
quality. Those subjective criteria and grade descriptors were given to the 15 main
panels and 67 sub-panels which were to make the judgements, with the invitation that
– where they considered it necessary – they amplified the definitions, for both the
criteria and the grades. Many accepted this invitation, and produced extended
definitions of either or both of the criteria and the grade descriptors. None, however,
was able to provide a verbal description that was any more exact than the generic set; they merely re-wrote the subjective criteria that they will deploy and how they will place items into the five grades.

This major exercise did very little to formalise the evaluation procedure, therefore. In many ways this is not surprising. Most academics work within the constraints of their training and socialisation as members of one or more academic communities which have shared appreciations of what their (sub-)discipline studies, what its goals are, and how they pursue original research. They practice normal science – either applying existing methodologies (of a technical nature in some, scientific, disciplines) to address unsolved problems in a linear progression or using accepted means to broaden and/or deepen understanding of the human condition. Occasionally they will do something particularly original – solve a very difficult technical or analytical problem, for example, or undertake a major synthesis of existing knowledge, or explore a new theme with novel methods – and have a substantial impact accordingly. But these ‘paradigm shifters’ are (hopefully) rare – otherwise the nature of a discipline will be frequently overturned, making it very difficult to identify consistent criteria on which to judge its contents. Furthermore, the full impact of any ‘revolutionary’ piece may only be appreciated some years after the work has been published so that its assessment by an RAE sub-panel is based on evaluation of its potential impact only, which is clearly a highly subjective task.

Given this description of the nature of academic research activity, the implication to be drawn with regard to the RAE assessments is that a majority should only be allocated a 2* grade, which most of the descriptions discussed here suggest is the category for the great bulk of normal science: the occasional highly original piece of work in that mould, plus some paradigm-shifting pieces at the exemplar scale, would be allocated 3*. The accolade of 4* should be extremely rare: most sub-panels expand on the term ‘world-leading’ in the generic descriptors using terminology suggesting pieces of work that have very substantial impact on all workers in the field in which they are published (which itself is difficult to define). Such a limited number of 4* grades is however extremely unlikely – for political rather than intellectual reasons. If a sub-panel is grudging with its 4* (and probably 3*) grades it will, in effect, be saying that there is little work done in the UK in that discipline which is world-leading, which will place the discipline at a disadvantage – not least in the allocation of funding – relative to others where more high grades are allocated.

A clear response to this problem was recommended to, but rejected by, the Funding Councils in the Roberts Report. It proposed a threefold classification of outputs, and recommended that ‘panels be given guidelines on expected proportions of one star, two star and three star ratings. These proportions should be the same for each unit of assessment. If a panel awarded grades which were more or less generous than anticipated in the guidelines, these grades would have to be confirmed through moderation’ (Roberts, 2003: Recommendation 5c, #171). In the 2001 RAE panels were required to draw up statements describing their working methods and assessment criteria. Roberts proposed retention of many of the key features of the 2001 exercise, including that ‘panels establish their own assessment criteria in consultation with their research community’ (Roberts, 2003, #12f), adding the specific recommendation that ‘panels should ensure that their criteria statements enable them to guarantee that practice-based and applied research are assessed
according to criteria which reflect the characteristics of excellence in those types of research in those disciplines’ (Roberts, 2003: Recommendation 7b, #204).

If the Funding Councils had accepted Roberts’ recommendations, the panels and sub-panels would have faced a very different task. They would have been invited to make relative not absolute judgements – if 20 per cent of outputs were to get the top grade, for example, then they would have had to rank order them on the agreed criteria. But instead they were asked to make absolute judgements – not that this piece of work is among the best 20 per cent produced in that discipline but that it meets standards of originality, rigour and significance which mean that, for example, every worker in that field will see it as making a major contribution to the field’s advancement. Both types of judgement are difficult, but absolute judgements are more difficult than relative ones, especially when the criteria for making such subjective judgements are – necessarily? – vague and difficult to condense into any more rigorous form (although, as indicated above, there are implications that different types of metric could be substituted for the subjective judgements in certain circumstances). As demonstrated here, the RAE2008 panels and sub-panels were unable to expand on the criteria and extend the generic grade descriptors in any way that enhanced the outsider’s appreciation of what would be done – let alone gave the sub-panel members any clearer guidance on how to undertake their task.

All of the UK RAEs have been exercises in subjective judgement by experts applying their professional expertise, experience and knowledge to try and identify where the best research is being done. They have differed in degree but not in kind from the wide range of other subjective judgement processes undertaken in academic life – from marking student essays to assessing a research grant application or journal article submission. There were attempts – as in the 2001 exercise – to systematise these subjective judgements with grade descriptors, but these were at best vague and at worst uninformative: the experts simply rank ordered UoAs and then drew dividing lines between the grades where they considered it appropriate. For the 2008 RAE, the Funding Councils have sought to systematise the grading procedure even more – in part because they are now asking panels and sub-panels to evaluate separately every piece of research output submitted and to build a quality profile for each UoA based on evaluations of its submitted research outputs, description of its research environment, and details of its members’ esteem indicators. To advance that, they identified three explicit criteria against which the judgements were to be made, as the basis for placing each output in one of five quality categories, themselves given descriptive statements encapsulating the quality of work. The panels and sub-panels were then invited to consult with their academic communities and, where they considered it necessary, provide further explication of the criteria and/or the grade descriptors, to aid their decision-making and enhance the transparency, credibility and legitimacy of the exercise. Most did, though to little effect: they have done little more than rewrite one description of subjective decision-making into a different set of words and have done nothing to make the procedure either easier for those who undertake it or more transparent to those subject to it.

The RAEs have come in for a great deal of criticism within the UK academic community – which hasn’t prevented the model being adopted and adapted elsewhere (see, for example, Worthington and Hodgson, 2006; Johnston, 2006c). RAEs, as the Roberts report recognised and was reflected in its consultations, are essentially
exercises dependent ‘upon expert peer review to identify the best research’, comparable in detail if not in scale to many other such exercises in which academics are involved. Experts are asked to classify materials into categories based on their perceived quality. If they are to do that in a relative sense, they need little more guidance other than a statement of the expected proportions in each category. But if they are to make absolute judgements, then they need clear criteria and guidelines. The experience of the build-up to RAE2008 is that this is not feasible – all that can be produced is a series of subjective statements because precise benchmarking is not possible. Academic excellence, Justice Potter Stewart might well have decided if the RAE had been contested before him in the US Supreme Court, can be identified but not defined.

Acknowledgements

Thanks to Graham Badley, Rita Johnston and Kelvyn Jones for valuable comments on drafts of this essay.

Notes

1 Though an intermediate stage of ‘revise and resubmit’ is often used, thereby delaying the final decision until further material is available.
2 The source is in the cases of *Jacobellis v Ohio* (378 U.S. 184, 1964) and the full text is: ‘I shall not today attempt further to define the kinds of material I understand to be embraced within that shorthand description [i.e. hard-core pornography]; and perhaps I could never succeed in intelligibly doing so. But I know it when I see it…’.  
3 Increasingly over the six RAEs, the Units of Assessment have become less closely linked to University departments. The latter can choose to which UoA they submit either all or part of their activity/staff: a School of Mathematics may split its work between as many as three, if not four, UoAs (20-23: see Table 1), for example, and a Department of Geography might choose to submit its physical geography/ers to the Earth Sciences UoA – either because it considers this will produce a better evaluation of the work than if it was submitted to the Geography and Environmental Studies UoA or because they thought they would benefit more financially – or both. Thus UoAs cover types of work (in most cases defined substantively), which increasingly – given the many restructurings of recent decade- cannot be directly mapped into university departmental structures.
4 For details on the latest formula see http://www.hefce.ac.uk/research/funding/qrfunding/.
5 Only one panel (E) split on this issue: UoA 17 decided to weight outputs at 65 per cent and UoAs 18-19 at 60. The latter, along with panel G, also gave the highest weightings to esteem indicators (15, 20 and 30 per cent respectively), with all of the others allocating either 5 or 10.
6 Four separate Funding Councils are responsible for UK universities – Higher Education Funding Council for England (HEFCE); Scottish Funding Council; Higher Education Funding Council for Wales; Department for Employment and Learning (Northern Ireland). They combine to conduct the RAEs, which are managed from the HEFCE offices.
7 The terminology was not new. At the previous RAE (2001) UoAs were assessed on a 7-point scale, with individual levels being specified in terms such as ‘Quality that equates to attainable levels of international excellence in more than half of the research activity submitted and attainable levels of national excellence in the remainder’ (this for the highest – 5* – grade). The full set of grade descriptors is in Roberts (2003, Appendix C).
8 This was, of course, an ironic statement since Newton made it when writing to Robert Hooke – who was very small. The phrase dates back to St Bernard of Chartres in c.1130.
9 This is illustrated by the growth of ‘string theory’ in physics over recent decades.
10 The issue of disciplines (or, at least, disciplinary RAE panels and sub-panels) having lists of journals according to their perceived quality, so that publication in some is seen as much more prestigious and indicative of excellent than in others, has long concerned some subject communities (in most of which there is clear evidence that academics select their ‘four best outputs’ from a relatively limited range: Johnston, 2003). Thus, for example, the sub-panel for UoA 39 (Politics and international studies) in RAE2008 not only stated that it ‘will not establish a list of the relative standing of journals. It
recognises that some types of research are published in less prominent or more specialised journals’ but also that it ‘will not establish a list of the relative standing of publishers’ for books and monographs because ‘some types of research are published by less prominent or more specialist publishers’. Law (UoA 38) made a similar statement with respect to journals

Sub-panels are able to send work to other sub-panels for evaluation and request their specialist advice, but whether this will benefit the work so referred is a moot point.

A few extend the range: European Studies (UoA 50), for example, refers to ‘quality, ambition and originality of the research design and methodology’ (my emphasis).

This concern with plurality of competing approaches is a particular concern of (human) geographers and sociologists. For the latter UoA 41’s document states that ‘the sub-panel will use these criteria to assess all forms of research. To ensure that new and emerging areas are not disadvantaged, the benchmarks of excellence will be applied flexibly when assessing research at the cutting edge of the research area. In view of the diverse nature of the discipline of sociology, the sub-panel understands the quality descriptors to relate to indicators within fields, sub-fields and cognate areas…’.

The Architecture sub-panel (UoA30) defines rigour as including ‘research processes which are not necessarily systematic or linear, yet demonstrate intellectual precision and material integrity, and innovations in process and/or product in relation to its context’.

Similar phrases were used in the 2001 RAE. For example, the History panel referred to ‘national excellence’ as embracing ‘highly competent work within existing paradigms which does not necessarily alter interpretations’ and the Music panel defined ‘internationally excellent work’ as ‘likely to contribute to the formation of new paradigms or the significant extension of existing ones’. Other definitions also draw on the 2001 documentation: the Geography panel, for example, understood ‘research of international excellence to be that which is judged to shape the development of the field with respect to new empirical findings, or new concepts, or new methodologies, or new applications, and is expected to steer research agendas for the foreseeable future’. (All of these quotes are taken from http://www.hero.ac.uk/rae/Pubs/5_99/RAE5_99.doc .

Panel E includes Physics which discipline, according to two commentators, has been taken over by a single paradigm (string theory) in recent decades to the exclusion of almost all else (Smolin, 2006)!

The fifth category – unclassified – is ignored here: there are no substantial variations in its definition.

Forty years ago J. D. Bernal said that ‘there is an enormous increase in the amount of publication each year, and although you may be able to produce 50 per cent more papers per annum, it is doubtful whether you could read them. It is also doubtful whether there is any use reading them; it is really a sorting process. If you read one in a hundred or one in a thousand of the papers you will find out the state of the field …’ (quoted in Brown, 2005: 451).

References


### TABLE 1.
The panels and subpanels

<table>
<thead>
<tr>
<th>Panel</th>
<th>Sub-panels</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>34. Economics and econometrics; 35. Accounting and finance; 36. Business and management studies; 37. Library and information management.</td>
</tr>
<tr>
<td>J</td>
<td>38. Law; 39. Politics and international studies; 40. Social work and social policy &amp; administration; 41. Sociology; 42. Anthropology; 43. Development studies.</td>
</tr>
<tr>
<td>K</td>
<td>44. Psychology; 45. Education; 46. Sports-related studies.</td>
</tr>
<tr>
<td>L</td>
<td>47. American studies and Anglophone area studies; 48. Middle Eastern and African studies; 49. Asian studies; 50. European studies.</td>
</tr>
<tr>
<td>N</td>
<td>59. Classics, ancient history, Byzantine and modern Greek studies; 60. Philosophy; 61. Theology, divinity and religious studies; 62. History.</td>
</tr>
<tr>
<td>O</td>
<td>63. Art and design; 64. History of art, architecture and design; 65. Drama, dance and performing arts; 66. Communication, cultural and media studies; 67. Music.</td>
</tr>
</tbody>
</table>
**TABLE 2**

Definitions of quality levels

<table>
<thead>
<tr>
<th>Star</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Four star</td>
<td>Quality that is world-leading in terms of originality, significance and rigour.</td>
</tr>
<tr>
<td>Three star</td>
<td>Quality that is internationally excellent in terms of originality, significance and rigour but which nonetheless falls short of the highest standards of excellence.</td>
</tr>
<tr>
<td>Two star</td>
<td>Quality that is recognised internationally in terms of originality, significance and rigour.</td>
</tr>
<tr>
<td>One star</td>
<td>Quality that is recognised nationally in terms of originality, significance and rigour.</td>
</tr>
<tr>
<td>Unclassified</td>
<td>Quality that falls below the standard of nationally recognised work. Or work which does not meet the published definition of research for the purposes of this assessment.</td>
</tr>
</tbody>
</table>
APPENDIX I
Definitions of originality, significance and rigour.

UoAs 34, 35, 36, 37
Originality in terms, for example, of innovation or distinctiveness in the methodological approach or in the datasets used, of research question posed, or of the underlying hypothesis or theoretical frameworks applied. Significance in terms, for example, of providing understanding of economic behaviour, or insight into the construction of economic policy making, or in the development of the discipline of economics and econometrics. Rigour in terms, for example, of the contextualisation of the work, the strength, appropriateness and intellectual coherence of approach, or the extent to which this supports the research outcomes.

UoA 39
Originality, recognising that new knowledge and understanding can come in a wide variety of forms. Rigour, recognising the need for sensitivity to a plurality of approaches and methods while endorsing high standards of scholarship. Significance, recognising the different needs and expectations of academic and other communities, giving due weight to potential as well as actual significance.

UoA 41
The quality of outputs will be assessed in relation to the extent that work demonstrates originality, rigour and significance. In terms of originality, the sub-panel will consider the intellectually innovative character of the research to be the most important characteristic. In terms of rigour, the sub-panel will consider the robustness of the argument and methodology, and the appropriateness of method(s). In terms of significance, the sub-panel will consider the contribution of the output towards setting or advancing the intellectual or policy agenda or generating new audiences; the extent to which the research activity is generally recognised as being of high quality in and beyond the subject community or specialism; the degree to which the output has promoted the discipline, raised its profile among other disciplines, and developed interdisciplinary research; and the effect and influence generated in user communities by the output.

UoA 42
In all cases the sub-panel will expect originality, significance and rigour to be embodied in work that presents new empirical material, discoveries or substantive findings, generates novel conceptual or theoretical syntheses, tests hypotheses, and/or pioneers innovative research agendas and directions.

UoA 43
In terms of originality, the sub-panel will consider the contribution to the advancement of theory and/or the innovative character of the research (ie, new and significant research questions, new kinds of evidence, new interdisciplinary and multidisciplinary insights). In terms of rigour, the sub-panel’s criteria for excellence will be theoretical or conceptual rigour; the clarity of research and its relation to research design and practical methodology; advances in analytical methods; and the interpretation of evidence which sets new standards for research in the field. In terms of significance, the sub-panel will consider the prescriptive and/or critical contributions of the output towards the advancement of policy agendas and policy
practice which affect developing countries; and the effect and influence generated by the theoretical or empirical insights of the research among academic and other users.

**UoAs 44, 45, 46**

*Originality* is a characteristic of research which is not merely a replication of other work or simply applies well-used methods to straightforward problems, but which engages with new or complex problems or debates and/or tackles existing problems in new ways. So, for example, a review of existing research can demonstrate originality if it analyses and/or syntheses the field in new ways, providing new and salient conceptualisations. Originality can also lie on the development of innovative designs, methods and methodologies, analytical models or theories and concepts.

*Significance* can be judged in different ways according to whether the research is basic, strategic or applied. Research has, or has the potential to have, considerable significance if it breaks new theoretical or methodological ground, provides new social science knowledge or tackles important practical, current problems, and provides trustworthy results in some field of education. These results might be empirical or analytical and theoretical, providing new (and sometimes challenging) conceptualisations, and evidence for audiences ranging from academics to policymakers and practitioners. Ways of evaluating the significance of research include judging its effect or potential effect on the development of the field, examining contributions to existing debates, and assessing its impact or potential impact on policy and practice. The nature and degree of immediate impact on policymakers or practitioners will provide some useful indication of significance in terms of ‘value for use’. However, there may be reasons for high impact that are not dependent on research quality; and, equally, in many cases the observable impact of high quality research is achieved only over the longer term. Theoretical and more analytical research can also be of high significance if it takes forward the state of current international knowledge in its field, and has influenced, or has the potential to influence, the work of other theoreticians....

**Rigour** can be judged in many ways, and can helpfully be associated with methodological and theoretical robustness and the use of a systematic approach. It includes traditional qualities such as reliability and validity, and also qualities such as integrity, consistency of argument and consideration of ethnical issues. It certainly entails demonstrating a sound background of scholarship, in the sense of familiarity and engagement with relevant literature, both substantive and methodological. Different dimensions of rigour will be important in different types of research but rigour can best be assessed on a case by case basis using whichever dimensions are most appropriate. In the case of outputs that are primarily directed towards utility, it is still the rigour of the underpinning research work that will be assessed and will need to be evident.

**Panel L**

*Originality*

Imaginative range

*Significance*, as demonstrated by the extent to which knowledge or understanding in the field has been increased or practice has been or is likely to be improved.

*Rigour*

**UoAs 51-58**

*Originality*: an intellectual advance or an important and innovative contribution to understanding and knowledge. This may include substantive empirical findings, new interpretations or insights, development of new theoretical frameworks and conceptual models, and innovative methodologies.
Significance: imaginative scope; importance of the issues addressed; impact or implications for other researchers and users.

Rigour: intellectual coherence, methodological precision and analytical power; accuracy and depth of scholarship; evidence of awareness of and appropriate

UoA 62

Originality can be in the form of reshaping interpretations or approaches or opening up new sources, new data or material. Significance will be judged on the basis of, for example, depth and likely lasting scholarly value. Rigour will be judged on, for example, accuracy, clarity and standards of scholarship.

Panel O

Significance
The degree to which work has enhanced, or is likely to enhance, knowledge, thinking, understanding and/or practice in its field.

Originality
The degree to which the work has developed new formulations or data and/or initiated new methods and/or forms of expression.

Rigour
The degree of intellectual precision and/or systematic method and/or integrity embodied in the research.