Research Excellence Framework

Higher Education Funding Council for England

February 2008

Executive Summary

This response has been prepared following a number of meetings of Fellows of the Royal Academy of Engineering, consultation with a wider group of Academy Fellows and a joint meeting with HEFCE to discuss potentially suitable metrics for the assessment of excellence in engineering research.

The Royal Academy of Engineering has taken a strong interest in the quality of engineering research over many years both as a funder of research posts within the UK university system and as the UK academy for engineering. The Academy's work in the area includes *Measuring Excellence in Engineering Research*[1], *The Assessment of Research Quality in Engineering Disciplines*[2], *The Future of Engineering Research*[3] and, in conjunction with EPSRC, *The Wealth of a Nation – An Evaluation of Engineering Research in the United Kingdom*[4].

In the January 2000 report, *Measuring Excellence in Engineering Research*, the Academy used the terms Mode 1 to describe engineering research focusing on the creation of new scientific knowledge and Mode 2 to describe engineering research impacting on society through wealth creation. For convenience and consistency, this response refers to these modes of research to describe mostly academic research for the pursuit of knowledge and research with more direct industrial application.

> From *Measuring Excellence in Engineering Research*, RAEng, 2000.
>
> **Mode 1 Research**: focusing on the creation of new scientific knowledge needed to underpin novel and innovative engineering devices and processes. Here attention is normally focused on single-disciplinary problems which are "set and solved…by largely academic community, and are communicated through institutional channels"[5] (e.g. publications) and driven by curiosity and the desire for new insights. A country with an engineering base that is weak in the production of Mode 1 research will be relatively ill-placed to develop the radical innovations that spring from new scientific knowledge.
>
> **Mode 2 Research**: impacting on society through enhancement of wealth creation and quality of life by producing knowledge in the context of application in non-academic and academic communities – activities which may be single of multidisciplinary in nature. Mode 2 research manifests itself ultimately through people centred activities, (for example, the provision of advice, the appearance of new products, processes, start-up companies and collaborative networks) and scholarship. The diffusion of the results of Mode 2 research occurs as it is produced: its "outputs" are both difficult to identify or separate from the production process. A lack of Mode 2 research will fundamentally undermine the strength of the research base, not simply in the application of new knowledge but also in the production of new research. Academic excellence and relevance to the user group are essential partners in Mode 2 research.

[1] http://www.raeng.org.uk/news/publications/list/reports/Measuring_Excellence.pdf
[2] http://www.raeng.org.uk/news/publications/list/reports/Assessment_of_Research.pdf
[3] http://www.raeng.org.uk/news/publications/list/reports/Future_of_Engineering.pdf
[4] http://www.raeng.org.uk/news/publications/list/reports/EPSRC_Showcase.pdf
[5] Gibbons,M., Limoges,C., Nowotny, H., Schwartsmann,S., Scott, P., and Trow, M., *The New Production of Knowledge*, Sage, London, 1994

On the issue of metrics-based assessment of research, the Academy has also responded to the *Science and Innovation Investment Framework 2004-2014: Next Steps* HMT consultation[6] of June 2006 and the DfES consultation[7] *Reform of Higher Education Research Assessment and Funding* of September 2006. Both responses expressed concerns about the impact on engineering research of assessment systems based predominantly on citation analysis or research funding.

The Research Assessment Exercise has been the method used for determining the allocation of block, quality related, funding to universities for two decades. Whilst there is general agreement that it has increased the UK's research performance and international standing over that period, it has also suffered from a significant amount of 'mission creep'. The detail of analysis which awards star ratings to individual research departments has come to have many uses beyond the allocation of funding and could be argued to be important to a university's esteem and international standing albeit that such detail is unnecessary for the purposes of block grant allocation.

The objective of producing a radically simplified and more fit for purpose assessment system is applauded, however, the Academy believes there are significant problems with the current proposals for a metrics-based system and has come to the view that the currently proposed Research Assessment Framework would have dramatic effects on the way in which engineering research is carried out in the UK.

In particular, the balance of Mode 1 and Mode 2 research is under threat. The publication habits of Mode 1 and Mode 2 researchers differ greatly, with Mode 2 research output being much less amenable to citation analysis using conventional databases such as Web of Science. This reflects the recognition that engineering research is necessarily done within a context of application which extends well beyond the academic peer community.

The basket of metrics proposed in the Research Excellence Framework appears to have been designed to measure academic excellence but fails to recognise or value impact in terms of useful output to industry, jobs created or contribution to the UK economy. This appears to run contrary to Government's determined efforts to improve the UK's innovation performance and to strengthen the role of universities in support of national innovation policy.

In addition to concerns about the metrics based system proposed, the Academy does not believe that science and engineering research should be assessed by a system which is separate from that applied to the arts, the humanities, and mathematics. Engineering research is increasingly multidisciplinary, including collaboration with the social sciences and particularly mathematicians. If it is unclear how such collaboration will influence income, it is likely to decline. Ideally, the Academy would suggest, all research assessment should include a light touch peer review.

[6] http://www.raeng.org.uk/policy/responses/pdf/Next_Steps_Response.pdf
[7] http://www.raeng.org.uk/policy/responses/pdf/Higher_Education_Research_Assessment.pdf

Question 1a: Do you endorse our proposals for defining the board group of science-based disciplines, and for dividing this into six main subject groups, in the context of our new approach to assessment and funding?
Question 1b: Are there issues in relation to specific disciplines within this framework that we should consider?

Whilst there is an undoubted need to simplify the way in which research excellence is assessed, leading to a higher level of aggregation within subject areas, boundary considerations remain an issue. In engineering research, it is increasingly common to work in multidisciplinary ways that cross subject boundaries. This introduces uncertainties as to how certain areas of research contribute to the overall funding of universities.

Emerging fields of engineering research, such as biomedical engineering, and established fields such as chemical engineering will find themselves assessed, in part, as either biological sciences or physical sciences. The cultural differences across research boundaries and the differences in weighting that expert groups choose to apply to the basket of metrics will have implications for the balance of the research conducted and the publication habits of researchers.

Of more concern are areas such as computer science where multidisciplinary research crosses the boundary with mathematics or where more general applied engineering research has interactions with economics or social sciences. The implications for multidisciplinary research which crosses the divide between pure metrics and light touch peer review assessment systems is more uncertain.

Q2a: Do you agree that bibliometric indicators produced on the basis that we propose can provide a robust quality indicator in the context of our framework?
Q2b: Are there particular issues of significance needing to be resolved that we have not highlighted?

The Academy does not accept that the use of bibliometric indicators as proposed is capable of providing a robust indication of research quality for engineering subjects either generally or in the context of the proposed framework.

Research within engineering can be characterised as either Mode 1 or Mode 2. Mode 1 research is concerned with the advancement of human knowledge and is essentially similar to pure academic research in any of the other sciences. Mode 2 research is concerned with problem solving in a context of application and often results in useful processes or technologies with potential benefits for society or industry. These two types of research exist in all sectors of engineering.

Mode 2 engineering research is concerned with problems and processes. The result is that the output of this research is often developed in industry resulting in products and services which contribute directly to the economic wellbeing of the UK. This impact is not recognised in any bibliometric assessment because it is significantly less cited in academic journals. Another area where Mode 2 research can make a real impact is in the development of standards, which are used on a daily basis by industry but, again, rarely cited.

If the performance of a university, and hence its income from Quality Related block grants, is judged primarily on bibliometric indicators, it is likely that the balance of engineering research will shift away Mode 2 research, to the direct detriment of innovation and the UK economy.

Within engineering, the publication habits of researchers vary markedly from sector to sector. Chemical engineering has an adequate coverage in the Thompson Web of Science, allowing worthwhile citation analysis to be carried out. Most other fields of engineering research have poor coverage within the Web of Science. In order for citation analysis to be robust for engineering it is necessary to consider a wider set of publications and publication types, none of which are currently encompassed within Web of Science. The auditing of these larger groupings of journals and proceedings will be problematic but necessary, with a certain amount of judgment required on the part of expert groups as to the value of those citations that occur outside the Web of Science. It is possible that the Web of Science will expand its coverage to include all that is needed to fairly assess engineering research in future, allowing this to be achieved at significantly lower cost and on an equal basis to other sciences, but it cannot be assumed.

The complexity and cost of auditing an expanded citation analysis could well outweigh the projected savings in adopting a bibliometrics based system.

Q3a: What are the key issues that we should consider in developing light touch peer review for non-science-based disciplines?
Q3b: What are the main options for the form and conduct of this review?

As explained above, the Academy is firmly of the opinion that separate assessment methods for science and engineering, and others, is a flawed concept which will lead to uncertainties at the boundaries.

However the light touch peer review process develops, it must implicitly value collaborations across the arts and humanities / sciences boundary. Indeed, this may be significantly easier to achieve than ensuring that it is valued by the metrics driven assessment on the science side.

Some of the characteristics of impact that fail to be captured in the metrics based assessment of engineering research also apply to fields such as economics, management, architecture and design. The proposed light touch peer review should be able to value this impact in a way that cannot be done for the sciences under the proposed system.

Q4: 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?

The Academy held a one day meeting and workshop in January 2008 to explore the issue of additional quantitative information that could be used to capture impact beyond academic excellence. Four workshops addressed the problem from slightly differing perspectives and answered a related set of questions. The workshops were as below:-

Workshop 1 – characteristics of excellence

- What are the characteristics of excellent engineering research?
- Are these adequately (for the purposes of allocating block grants) reflected in the proposed advanced bibliometrics, research income and research student numbers metrics?
- Which characteristics could be used to form the basis of suitable metrics?

Workshop 2 – behavioural changes

- Is there scope for the proposed new system to encourage “game playing” and behavioural changes within HEIs detrimental to improving excellence in engineering research?
- What changes might be made to the proposed system to combat such behavioural changes?
- What additional metrics might be used that could ameliorate such behavioural changes given suitable weighting?

Workshop 3 – applied and basic research

- Should the Research Excellence Framework be implemented as currently envisaged by HEFCE, how would it impact on engineering research, particularly the split of Mode 1 versus Mode 2 research?
- How important is it in the overall REF process, given the levels of aggregation that it will operate to, that excellence in Mode 2 research is specifically measured?
- Can Mode 2 research that has lower citation levels be valued by any other measure?

Workshop 4 – industrial perspective

- As potential end users and sponsors of engineering research, what is the industrial view of excellence in engineering research?
- Is this captured in the metric proposed for the REF? Could research income be a suitable metric?
- How can user value be captured and by what potential metric?

A number of ideas were generated from the workshops for metrics which could supplement the use of citation analysis, research income and PhD numbers, however, in the short time available, only the seeds of ideas were developed. If any of the ideas were to be taken forward, it would be necessary to develop and research them to the same level as has been done by the Centre of Science and Technology Studies (CWST) at Leiden University for citation analysis. This is no trivial task and may delay the introduction of the Research Excellence Framework.

Notes of the meeting are attached as Annex A and the output from the workshops is detailed below.

Workshop 1 – characteristics of excellence

Characteristics of excellent engineering research were seen as the take-up of ideas by industry (including contribution to practice), advancement of knowledge, PhD training and international Reputation. Advancement of knowledge was adequately reflected in the Research Assessment Framework, but the others were not.

Advancement of knowledge was amenable to measurement through a good publications record, but was not necessarily reflected by the level of research grants attracted or the number of students in a department. Bibliometrics measures were thought to be only an indicator of knowledge advancement rather than a direct measure. Contribution to practice was felt not to be reflected in bibliometrics measures or student numbers. There was thought to be some correlation with research income, but only income from users rather than Research Councils.

International recognition could in part be measured by bibliometrics measures, but not by student numbers or research income.

Other metrics that could reflect these were discussed.

- Advancing knowledge might be reflected by academic prizes, key note addresses at conferences, international collaborations and membership of international technical committees.

- Contribution to practice might be reflected by membership of standards committees, consultancies held, number of spin-outs, involvement in knowledge transfer partnerships, numbers or value of patents and intellectual property, collaboration directly with users and participation in user seminars.

- International recognition might be reflected in chairing of conferences and programme committees, invited presentations and keynotes at international conferences, elected positions on international bodies, VP positions, international prizes and international consultancies or projects.

Workshop 2 – behavioural changes

Danger of significant behavioral changes was seen in the imposition of the basket of metrics as currently envisaged in the Research Excellence Framework. Of particular concern was the possibility of citation rings evolving in certain research areas, a mitigation against participation on conferences or international exchanges and a move towards more Mode 1 research, away from Mode 2 research.

In order to mitigate against these behavioral changes it would be necessary to adopt a more inclusive database for citation analysis, maintain a low weighting for citation analysis and introduce an element of esteem into the assessment (which could probably only be achieved by a light touch peer review).

Other metrics to address these behavioral changes included:

- Rating industrial income much more highly than Research Council income.

- Numbers of PhD's graduated and the number of man-months of post-doctoral work conducted

- Esteem (probably qualitative rather than quantitative).

Workshop 3 – applied and basic research

The workshop's view was that the Research Excellence Framework should not be implemented in its current form. A formulaic approach to assessing excellence did not seem able to take into account enough variables. The level of development of citation analysis seemed to be poor and still at a very experimental basis (particularly in terms of its impact on real systems). Given the immaturity of citation analysis, a peer review system was needed for balance.

Even at the proposed levels of aggregation, it was felt to be important to have a rigorous measure of the quality of Mode 2 research as well as Mode 1 research. The main reason for this was the temptation for researchers to switch from Mode 2 to Mode 1 for better citation results.

Metrics that could reflect Mode 2 research better included:-

- The number of facilities within an institution (eg industrial scale labs)
- Number or value of spin-outs
- Strategic links
- Number of trained researchers moving into industry
- Industrial R&D income
- Patents registered
- Number of jobs created.

Workshop 4 – industrial perspective

Industry/academic collaborations are not easy and difficult to recognise or value in the current RAE system. Any new proposed system must improve on this. The nature of the relationship between universities and industries is complex and multi-levelled involving two way flows of information and people. Industry also values collaboration as method of bolstering the pipeline of skilled graduates available for recruitment.

The group felt that none of this user value was captured by the proposed framework and the following indicators provided some insight into it:-

- The gearing of the research income between industry and Research Councils. It was felt that the relative weighting for research income to citation analysis should be 3:1.
- The length of relationships between universities and industrial partners
- The number of skilled staff moving between industry and academia.

The Academy would be pleased to work with HEFCE to develop any of these indicators further.

Q5: 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 role of the expert groups appears to be to decide suitable weighting for the metrics being used in the assessment. As the Academy has argued that further indicators of quality and impact should be taken into account in the framework, it is essential that the expert groups be reconstituted to be able to properly assess the new indicators suggested.

For engineering research, it is of utmost importance that user groups, such as industry, be involved in the expert groups. User impact is likely to be difficult to capture as a purely quantitative metric and the industrial member of the expert group will need to be able to take a value judgment on the evidence available.

The role of international members of expert panels, while useful, seems to be superfluous. As the framework is only designed to allocate funding on merit between

UK universities, international comparisons bring a level of analysis which is unnecessary.

Q6: 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 minimize the burden as we introduce the new arrangements?

The implementation of two separate assessment systems for science and engineering, and arts and humanities will impose additional burdens on university administrations. Any projected easing of burdens through the use of metrics will be severely eroded.

It is not entirely clear as to what the effect might be on the level of applications for Research Council funding. If universities see a potential threat to funding streams from the implementation of the new framework, some are likely to attempt to secure additional funding through the Research Councils, putting additional stress on the peer review system. This should be avoided at all costs.

Q7: 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?

An over reliance on bibliometric indicators may be indirectly discriminatory as female researchers are more likely to have changed names during the course of their research careers and be more prone to citations being lost in the analysis. The auditing of citations to ensure that this does not become an issue will be particularly burdensome in the early years of the framework.

Q8: Do you have any further comments about our proposals which are not covered by the above questions?

The Academy is concerned that the framework as described cannot capture the impact or user value of Mode 2 engineering research. Whilst the Academy has suggested various indicators that might capture something of the user value and impact and has offered to work with HEFCE to further develop these, the overwhelming view of Academy Fellows is that these may never be adequately captured by metrics and some level of qualitative assessment is highly desirable.

The Academy also believes that, as the framework is introduced, it is important to preserve the level of QR funding that is awarded on the basis of the quality of engineering research and, as engineering demonstrates its direct links to the innovation performance of the UK, the funding for engineering should increase over time.

Submitted by:
Mr Philip Greenish CBE
Chief Executive
The Royal Academy of Engineering

Prepared by:
Richard Płoszek
Policy Advisor
14 February 2008

The Royal Academy of Engineering

HEFCE Metrics Meeting and Workshop

Notes of a meeting held by the Royal Academy of Engineering on 25 January 2008

1. Introduction

The Government asked HEFCE to develop a new framework for research assessment and funding that was founded on the use of quantitative information (metrics). In science and engineering this would displace the current RAE system and its significant element of peer review at the level of Units of Assessment. The planned new metrics-based approach would be introduced gradually between 2010 and 2014 following the 2008 RAE.

Based on its past activity in relation to the measurement of excellence in research, The Academy had some concerns about how the proposed metrics would affect engineering. A number of meetings were held with HEFCE to discuss the matter. In consequence, the Academy agreed to work with HEFCE to explore concerns and to propose additional measures that would better reflect excellence in engineering research.

In order to develop these additional measures and to build on the work already conducted by HEFCE, the Academy held a one-day seminar during which the HEFCE proposals were explained, the shortcomings for engineering disciplines were explored and possible additional metrics were suggested. Delegates participated in an afternoon workshop session, contributing to the design of metrics that better reflect quality of Engineering Research Quality.

Programme

10:00 Registration and Coffee

10.30 Introduction by Chair
Dr Sue Ion FREng

10.40 The Development of the Research Excellence Framework
Dr Rama Thirunamachandran, Director (Research and Knowledge Transfer), HEFCE

11.00 Bibliometrics as a Measure of Research Excellence
Dr Henk Moed, Centre for Science and Technology Studies (CWTS), Leiden University

11.40 Questions and discussion – HEFCE and Dr Moed team

12:05 A perspective from EPSRC
Prof David Delpy FREng

12:25 Lunch

13:30 Panel Discussion
– introduced by 5-minute presentations from invited speakers

14:30 Workshop Sessions

15:30 Reporting back and Discussion
introduced by rapporteurs from each workshop

16:00 Closing Remarks
Dr Sue Ion FREng

2. The Development of the Research Excellence Framework

Dr Rama Thirunamachandran began by presenting some general statistics on UK research compared to the Global situation, showing the UK in second place in terms of volume to the USA. Originally, the Research Assessment Exercise (RAE) was introduced in order to bring the UK up the ranking table and to compete better with the USA. The system of dual support funding, with a Quality Related (QR) block grant to HEIs to maintain and improve research facilities, had been successful in improving the UK's research performance.

Starting in 1986, the RAE has become progressively more complex, moving from what was originally a light touch peer review based on objective opinions, to a system with many checks and balances and a complex, but essentially transparent, procedure. Dr Thirunamachandran postulated that, although improvements had been seen from the introduction of the RAE, the cost and complexity of the system meant that the benefits were flattening out.

In the Pre-Budget report of March 2006, the then Chancellor, Gordon Brown, indicated that the Treasury, while confirming its commitment to the dual funding system, would like the RAE to move to a system based on easily gatherable metrics. The initial suggestion was that this could be achieved by basing the QR block grant allocation on the research income of HEIs. In the summer of 2006, the Department for Education and Skills consulted on a metrics based system of assessment based largely on using research income normalised against RAE results and concluded that subsequent assessments should be driven by metrics as far as is possible. HEFCE was then asked to develop a system based purely on metrics for STEM subjects but with a light touch peer review system for arts and humanities subjects. HEFCE issued a consultation on the proposals in November 2007 with a closing date of 14 February 2008.

The timetable for implementation of a new assessment system was given as driving all funding decisions by 2014 and informing funding from 2010 – 2011. In the interests of ensuring that RAE2008 ran smoothly, it was decided that RAE Panels would have no involvement in the development of metrics until the funding decisions for RAE 2008 had been made.

The aims of a new metrics based system were described as: to develop robust quality indicators for all disciplines to underpin our selective funding approach; to reduce burden; to avoid undesirable incentives; and to promote equal opportunities. In achieving these, it was also important not to revert to the pre-1986 system which allocated funding on the basis of the number of students enrolled.

As the purpose of the Research Assessment Framework is to allocate block grant to HEIs, it is not necessary to go into the level of detail that the RAE has done in the past. If the level of aggregation is raised from the current level of Units of Assessment (down at a department level) to a much higher subject grouping, this should allow a robust enough assessment at much lower cost. The errors introduced by taking a more aggregated view should, statistically, cancel one another out, maintaining the system's robustness.

The system proposed by HEFCE and detailed in the November 2007 consultation relies heavily on bibliometrics as part of the basket of metrics. Bibliometrics appears to be robust at the broad subject level, is based on advanced techniques (not using journal impact factors), is sensitive to disciplinary differences, benchmarked against international norms, transparent and replicable, and reduces burden and the scope for games playing. Expert groups are still required within the broad subject areas to fine tune the weighting of each metric.

Issues of importance to engineering research that have been identified since the start of the consultation, include the fact that the coverage of the Web of Science (the primary citation index used) is not as good as for other subjects and it is debatable as to whether the basket of metrics is able to capture any measure of value or impact. Dr Thirunamachandran invited the audience to suggest metrics that could address these shortcomings in a robust manner.

3. Bibliometrics as a Measure of Research Excellence

Dr Henk Moed of Leiden University described the work that he had done, commissioned by HEFCE, on the development of robust citation indexing. He first described the citation habits of various subject groups and how these can be normalised to provide a robust measure of the scientific excellence of a group of researchers. In particular, he described the two levels of assessment that the system needed to address; the policy level where large block grants were distributed at a university level; and the internal level where universities distributed those funds across schools and departments.

At various levels of analysis, the citation metric could be weighted so as not to disadvantage certain groups such as young researchers or emerging fields as opposed to established fields. The differing half life of papers from different fields in terms of citations in each year could also be taken into account in this way, but as the citation window proposed was five years, this was not seen as a significant issue.

Web of Science coverage for engineering subjects was shown to be lower than for average in the STEM subject groupings and the level of citation within Web of Science was also low. Dr Moed therefore conceded that it may be important to supplement the coverage in Web of Science with outside sources, possibly including conference proceedings; however, the citation analysis becomes increasingly complex if this is done.

Citations tended to be a good measure of how excellent the academic world thought a piece of work was, but this failed to capture technical advances to come out of engineering research. To address this, Dr Moed suggested that a form of patent analysis and of citations of patents could overcome this.

4. Questions and discussion – HEFCE and Leiden University

Q1. Why are standards not included in the bibliography? If an engineering researcher's work contributed to nation or international (ISO) standards, it would have lasting and real impact. Is there any way to include these documents in any citation analysis?

A1. It could be examined.

Q2. Because of the implementation of Full Economic Costing (FEC), there is not a pure separation of the two streams of the dual funding mechanism.

The nature of engineering research is that some is like physics which can be assessed under the proposed system and some is "learning by doing". In the latter class, citation practices differ markedly and researchers are more likely to cite original sources rather than the latest work in an area.

Q3. In the communications field, research is often problem based with strong industrial interaction. Why not design systems to encourage the behaviours that are seen as good, such as industrial interaction, rather than an emphasis on avoiding bad behaviors?

Although Dr Moad had talked about strategies for including patents and citations of patents in the bibliographic analysis, what about licensing agreements?

A3. Hopefully, part of the point of this meeting was to suggest such alternative metrics

Engineering has different publication strategies to other science subjects. In the proposed system it would be down to subject expert groups whether to include journals or conference proceeding in the citation analysis.

Q4. It seems clear that in any proposed assessment system that money will follow money. Have the effects of shifting money to highly cited research groups been explored?

A4. Yes – extra money would hopefully lead to increased performance and therefore more money from the next assessment, but the causality of the relationship is not well understood.

Q5. There is room for game playing in all of the proposed metrics – for example, if the number of PhD students is taken as a metric, universities could simply recruit more Doctoral students, but this could lead to reduction in the quality of PhDs awarded.

A5. HEFCE was aware of this sort of potential, but bearing in mind that only a certain number of PhD studentships were funded, there was little scope for a university to radically increase its numbers.

Q6. Could we simply go for a soft touch peer review for engineering?

A6. There is currently a debate as to where Computer Science should sit – it has many similar characteristics to Mathematics in terms of its publishing strategies.

Q7. Google Scholar appears to have far better coverage than Web of Science for engineering by a factor of 15. Why can't this be used for citation analysis?

A7. Google Scholar is not mature or robust enough.

Q8. In the current RAE, all the engineering panels have opted to use the lowest possible weighting for bibliometrics. Surely this will continue into the new system with expert groups choosing to use the lowest possible rating?

5. A perspective from EPSRC

Professor David Delpy FREng, opened by indicating that the Research Councils proposed to make a joint response to the HEFCE consultation and indicated that he was giving a research council view of the proposals.

The Research Assessment Exercise had, over the years of its existence, experienced significant mission drift. Its original purpose was to allocate block QR funding to HEIs, a task it is still used for. However, as its complexity had increased and the level of analysis increased, it had become a tool for the universities and Government to make judgements about research quality at a fairly fine level of aggregation. The loss of this information will have implications for academic management.

Comparing raw citation data to historical RAE ratings, there is a good correlation for chemical engineering, but a bad correlation for mechanical engineering. This means that the proposed metrics will reflect reality well in some sub-disciplines, but very badly in others. In particular, it is likely to undervalue work with impact on practice. Also, the nature of citation analysis means that any view derived from them will be retrospective, whereas RAE panels were able make a judgement on research that was underway at the time of the assessment and of HEI's future plans. With the implicit time lag compared to the RAE system, the new Research Excellence Framework is likely to provide a picture of the health of research five to ten years earlier.

There were good reasons for removing self citations from the analysis, but a downside of doing this is to devalue multidisciplinary research where researchers were often joint authors in cross disciplinary papers. Also, in new and emerging multidisciplinary areas, there was likely to be a far higher self citation due to a lack of other research groups in the area. The view of the Research Councils was that the split in assessment methods between STEM subjects and arts and humanities was a bad idea. Engineering research often now spans the boundary between STEM and Arts and Humanities, and separate assessment systems will confuse.

Although a case had been made that the proposed metrics system measured scientific quality of research, when considering engineering research, this was not a proxy for measuring impact. The only attempts to assess impact had been related to research income, which was also a poor proxy as it is, by nature, an input metric.

Any significant emphasis on citation analysis has the potential to change publication habits of researchers and this is a particular danger for some branches of engineering where Web of Science does not provide good coverage.

Although the level of aggregation proposed should be adequate for the allocation of block funding, an unintended consequence could be that universities with a world class research team in an otherwise more ordinary school would appear to fall down the international rankings with consequent implications for their ability to attract research sponsorship.

The final possible unintended consequence which was of particular concern to the Research Councils was the possibility of universities attempting to increase the proportion of funding through research grants and thus increasing the strain on the grant application process. Any proposed change in the assessment system should demonstrate an overall gain for the whole funding system.

6. Panel Discussion

Each panel member was given five minutes to give their personal thoughts or views on the proposed metrics and their implications. After this the floor was be opened to discussion.

- Prof Mike Sterling FREng, Vice Chancellor, University of Birmingham – advocating reduced administrative burden on HEIs without serious implications for the distribution of block grant. The granularity of the proposed system is fine enough to determine the level of block grant at an institution level. Are there opportunities in the proposed system to streamline HEI's administrative procedures and enhance research?

- Dr David Grant CBE FREng, Vice Chancellor, Cardiff University – bringing the view of an industrialist who has moved to academia, able to contrast the requirements of both. Will the lack of ratings on a department basis hinder HEI's attempts to market their services? Although VC's have great autonomy on how to distribute block grants, will the new system help or hinder them in this task?

- Prof Paul Cannon FREng, Chief Scientist, Communications Dept, QinetiQ – bringing the view of a research intensive organisation outside of academia with strong ties to the research community. Is the impact of engineering research properly captured in the new system so that funding flows to where impact as well as academic excellence is high? Will a shift in emphasis from impact to academic excellence in HEIs affect how organisations such as QinetiQ interact with them?

- Prof John Perkins CBE FREng, Dean of Engineering and Physical Sciences, University of Manchester – exploring the position of funders and how the changing system will affect the way funders consider where to fund research. The RAE has always given a fairly fine grained indication of research quality down to a departmental level which will be lost – how will this impact the way funders operate in the absence of this information?

6.1. Prof Mike Sterling FREng

Prof Sterling stated that based on past experience of two Research Assessment Exercises, he was broadly in favour of a metrics based approach so long as the weighting of metrics was appropriate and that the results were moderated against previous assessments to allow time for the system to adjust.

The purpose of the RAE was to allocate QR funding as a block grant to universities, but the QR funding is decreasing in value, being supplemented more by direct research grants.

He felt that the changes being proposed had their roots in politics rather than in the university sector and that as the drivers were political; detailed analysis from the point of the academic profession was of limited value.

While the research capacity is supported by the QR element of university funding, the long term growth of research capacity is more strongly linked to the subject choices of 15 year olds currently in school. This could be a better place to look to increase the capacity and quality of research.

Citation analysis needs to be just part of a basket of suitably weighted metrics. Income derived from Intellectual Property rights could be another metric used to broaden the base of the analysis.

Prof Sterling's most pressing concern with the proposed system was not the metrics themselves, but that it relied on a commercial database that was subject to the profit making motives of its owners. He wondered how sensible it was to have the analysis of the UK's research dependent on a commercial provider.

6.2. Dr David Grant CBE FREng

The Research Assessment Exercise had been good for Cardiff University, driving change and increasing quality and standing. Prof Grant was, however, very concerned about the cost of the RAE2008 which was taking up the time of 34 highly paid staff within the university. He was therefore in favour a simpler system that reduced administrative costs.

From the point of view of a university administrator, the prospect of running two separate assessment system (STEM and A&H) was not attractive and a light touch peer review system for all would be preferred.

Dr Grant was particularly concerned with perverse incentives created by the proposed new system, in particular the risk to the link between industry and academia. At a time when Government appeared to be promoting industry academia interaction as a driver of innovation, it seemed particularly perverse that a university funding system should not

reward it and possibly undermine it. He also questioned whether the social benefits that universities bring to their regions should not also be included in the assessment.

As far as timing was concerned, it seemed perverse that this consultation was proceeding at a time when the RAE2008 panels were in the process of deliberating the next round of funding. It seemed that this was the start of an experiment in university funding rather than the implementation of a robust system.

6.3. Prof Paul Cannon FREng

Prof Cannon expressed the view that having been involved in three previous RAEs, the system was crazy in terms of the administrative effort and cost to the institution. Any new system implemented must be simpler and cheaper.

As far as innovation was concerned, industry needed good relationships with academia built up over many years. These relationships depend on free exchanges of people as well as ideas, with industrialists spending time in universities, academics spending time in industry and a flow of competent graduates into the business. In this sort of environment, it is the applicability of research to an industry's needs that counts rather than the number of citations it generates.

Any proposed system to replace the RAE must take into account esteem and impact. Possibly indicators of impact might include development licences and patents generated. If citation analysis is to have a place, may be a much higher weighting should be given to industry citations (given their rareness) than academic to academic citations.

It seems increasingly common for Research Councils to pursue strategically important fields of research, encouraging bids in these areas. In order to make the system as joined-up as possible, the views of the Research Councils on the utility of a university to their objectives should be included in the assessment.

6.4. Prof John Perkins CBE FREng

Prof Perkins addressed the position of research funders. There was, he felt, an unimpeachable case for simplification of the system, but this must be achieved without behavioural consequences.

Key issues to funders concerning where to fund research were the quality of people, the quality of the research environment and the quality of research produced there. Good interactions between a university and industry are also of high importance.

In terms of assessing the quality of a research environment, it has been tempting in the past to use a department's RAE rating as a proxy, but this did not always give an accurate picture. In this sense, losing the fine grain aggregation of the RAE assessment system is no bad thing. As far as engineering research is concerned, the use of bibliometrics does not yield useful information.

An undesirable behavioural change caused by use of bibliometrics as an indicator can already be seen in the United States. The core of chemical engineering research is being eroded as researchers seek to publish papers that are suitable for chemistry journals rather than chemical engineering journals as the impact factor and citation rates for papers in the chemistry journals is higher.

7. Discussion

Q1. Much commercially sensitive industrially sponsored research is subject to confidentiality agreements. Would this cause problems with measuring income form

patents or licensing agreements? Would a qualitative view from industrialist still be required?

A1. Although not identifiable at a project level, this information is already accessible through public accounts.

It is certainly right for universities to be involved in links with industry and some sort of measure of the value of this link. However, there is a need to recognize and sustain the whole ladder of levels of industrial interaction and intellectual property value is just one rung. Additionally, IP income can be dominated by the value of just a few licencing agreements and income from intellectual property could be manipulated if used as a metric.

In terms of using research funding as a metric, it might be useful to split Research Council funding from industrial funding. A significantly higher weighting for industrial funding may ameliorate any rush to increase Research Council funding from in increase in the number of bids.

It was felt that that engineering research must have impact and consequently impact must be part of any assessment of research quality. It was noted that in some universities, the research with the greatest impact came from researches with an entrepreneurial spirit, but these researchers were often not prolific paper publishers. There was therefore a double threat to this category of high impact research is citation analysis has a high weighting in the analysis.

Regardless of the outcome of any review of research assessment systems, there was strong support for the proportion of QR funding given on the basis of engineering research to be preserved or even increased in successive assessments.

8. Workshops

Delegates were invited to take part on one of four workshops designed to suggest new metrics which could help measure excellence in engineering research, avoiding the problems previously identified with citation analysis. Each workshop addressed the problem from a slightly different perspective and answered a slightly different set of questions.

Workshop 1 – characteristics of excellence

- What are the characteristics of excellent engineering research?
- Are these adequately (for the purposes of allocating block grants) reflected in the proposed advanced bibliometrics, research income and research student numbers metrics?
- Which characteristics could be used to form the basis of suitable metrics?

Workshop 2 – behavioural changes

- Is there scope for the proposed new system to encourage “game playing” and behavioural changes within HEIs detrimental to improving excellence in engineering research?
- What changes might be made to the proposed system to combat such behavioural changes?
- What additional metrics might be used that could ameliorate such behavioural changes given suitable weighting?

Workshop 3 – applied and basic research

- Should the Research Excellence Framework be implemented as currently envisaged by HEFCE, how would it impact on engineering research, particularly the split of applied versus basic research?
- How important is it in the overall REF process, given the levels of aggregation that it will operate to, that excellence in applied research is specifically measured?
- Can applied research that has lower citation levels be valued by any other measure?

Workshop 4 – industrial perspective

- As potential end users and sponsors of engineering research, what is the industrial view of excellence in engineering research?
- Is this captured in the metric proposed for the REF? Could research income be a suitable metric?
- How can user value be captured and by what potential metric?

The format for the workshops was for each group to consider the first three questions in turn. After considering a question and coming up with a number of solutions, the whole workshop was invited to rate the answers in order of importance. The least supported solutions or answers were rejected and the groups then used the highest rated answers in their deliberation of the next question and so on.

8.1. Workshop 1 – characteristics of excellence

Characteristics of excellent engineering research were seen as the take-up of ideas by industry (including contribution to practice), advancement of knowledge, PhD training and international reputation. Advancement of knowledge was adequately reflected in the Research Assessment Framework, but the others were not.

Advancement of knowledge was amenable to measurement through a good publications record, but was not necessarily reflected by the level of research grants attracted or the number of students in a department. Bibliometrics measures were thought to be only an indicator of knowledge advancement rather than a direct measure. Contribution to practice was felt not to be reflected in bibliometrics measures or student numbers. There was thought to be some correlation with research income, but only income form users rather than Research Councils.

International recognition could in part be measured by bibliometrics measures, but not by student numbers or research income.

Other metrics that could reflect these were discussed.

- Advancing knowledge might be reflected by academic prizes, key note addresses at conferences, international collaborations and membership of international technical committees.

- Contribution to practice might be reflected by membership of standards committees, consultancies held, number of spin-outs, involvement in knowledge transfer partnerships, numbers or value of patents and intellectual property, collaboration directly with users and participation in user seminars.

- International recognition might be reflected in chairing of conferences and programme committees, invited presentations and keynotes at international conferences, elected positions on international bodies, VP positions, international prizes and international consultancies or projects.

8.2. Workshop 2 – behavioural changes

Danger of significant behavioral changes was seen in the imposition of the basket of metrics as currently envisaged in the Research Excellence Framework. Of particular concern was the possibility of citation rings evolving in certain research areas, a mitigation against participation in conferences or international exchanges and a move towards more pure (like science) research, away from applied research.

In order to mitigate against these behavioral changes it would be necessary to adopt a more inclusive database for citation analysis, maintain a low weighting for citation analysis and introduce an element of esteem into the assessment (which could probably only be achieved by a light touch peer review).

Other metrics to address these behavioral changes included:

- Rating industrial income much more highly than Research Council income.
- Numbers of PhD's graduated and the number of man-months of post-doctoral work conducted
- Esteem (probably qualitative rather than quantitative).

8.3. Workshop 3 – applied and basic research

The workshop's view was that the Research Excellence Framework should not be implemented in its current form. A formulaic approach to assessing excellence did not seem able to take into account enough variables. The level of development of citation analysis seemed to be poor and still at a very experimental basis (particularly in terms of its impact on real systems). Given the immaturity of citation analysis, a peer review system was needed for balance.

Even at the proposed levels of aggregation, it was felt to be important to have a rigorous measure of the quality of applied research as well as pure research. The main reason for this was the temptation for researchers to switch from applied to pure for better citation results.

Metrics that could reflect applied research better included:-

- The number of facilities within an institution (eg industrial scale labs)
- Number or value of spin-outs
- Strategic links
- Number of trained researchers moving into industry
- Industrial R&D income
- Patents registered
- Number of jobs created.

8.4. Workshop 4 – industrial perspective

Industry/academic collaborations are not easy and currently difficult to recognise or value in the current RAE system. Any new proposed system must improve on this. The nature of the relationship between universities and industries is complex and multi-levelled involving two way flows of information and people. Industry also values collaboration as method of bolstering the pipeline of skilled graduates available for recruitment.

- The group felt that none of this user value was captured by the proposed framework and the following indicators provided some insight into it:-

- The gearing of the research income between industry and Research Councils. It was felt that the relative weighting for research income to citation analysis should be 3:1.

- The length of relationships between universities and industrial partners

- The number of skilled staff moving between industry and academia.

9. Closing Remarks

In summing up the day's discussion, Dr Ion reflected that the meeting found the imposition of two separate assessment systems (for STEM and arts and Humanities) to be unacceptable. The imposition of the Research Assessment Framework as it stood would not enhance or encourage links between industry and academia, which seemed an odd outcome when other parts of Government were working hard to promote such links to improve innovation performance.

It was generally agreed that the system for assessing research excellence and allocating the block QR grant was in need of radical simplification but it seemed to be a bad idea to switch directly to what appeared to be a very experimental system. Some sort of trial was needed if any new system was to be seen as robust enough.

Lastly and possibly most significantly, the meeting urged that the amount of money going into the university system based on the assessment of engineering research should at least be preserved and probably increased.