THE EVALUATION OF RESEARCH: MOTIVES, METHODS AND MISUNDERSTANDINGS

Paper presented to the Research Group for Science, Technology and Innovation Studies (TaSTI), University of Tampere

June 2006

Professor John Taylor
Centre for Higher Education Management and Policy at Southampton (CHEMPaS)
University of Southampton
United Kingdom

jtaylor@soton.ac.uk

© Professor John Taylor
Introduction

Tracing the development of higher education in the United Kingdom over the past 20 years, a key factor has been the impact of the Research Assessment Exercise (RAE). The RAE raises intense feelings, from fierce condemnation to grudging acceptance to warm support. There is no doubt that the Exercise has had a profound effect on UK higher education, on the structure and organisation of universities, on funding, on human resources and on research output. The RAE has also had a wider, international impact, inspiring similar developments in, for example, Australia and New Zealand, but also influencing the evaluation of research in many other countries, including Ireland, the Netherlands and Portugal. The RAE therefore justifies detailed consideration by the researcher in higher education.

This paper, using the RAE as a framework for analysis, looks at the reasons why research is evaluated in higher education, both by governments and funding bodies, and within institutions and organisational units (motives); how this is undertaken, including the strengths and weaknesses of contrasting approaches (methods); and at some of the problems and issues that have emerged (misunderstandings). The paper is based upon a wider project looking at the impact of the RAE on UK higher education. The research involved detailed documentary analysis and a series of interviews with Vice-Chancellors and senior officers of universities; the research also draws upon the personal experience and observations of the author who worked as a higher education manager on the RAE (or its predecessor) in 1986, 1989, 1991, 1995 and 2001.

Motives

The funding of research in UK higher education institutions is based on a dual support system. Under this system, a number of Government funded Research Councils allocate grants to universities and to individual researchers to undertake specific research projects and programmes. Historically, these grants covered the direct costs of the research (specific equipment, travel, consumables, Research Assistants/Research Fellows, but NOT, except in very special circumstances, the cost of the researchers themselves or the indirect costs such as buildings, equipment infrastructure and support staff). The other side of the dual support system is Government funding for research allocated to institutions by the Higher Education Funding Councils in the form of a block grant, available for use at the discretion of the University. This stream of funding has traditionally funded the cost of researchers themselves and of the institutional infrastructure, including buildings and staff; this idea was often referred to as providing for the “well found laboratory”. It is this second stream of Government funding that is at stake when considering the development and progress of the RAE. (It should also be noted that the historical view of the dual support system is changing, with the Research Councils now moving to fund 80% and ultimately 100% of the total costs of research projects; the discretionary stream of funding remains in place to support research infrastructure and providing scope for universities to pursue their own strategies for research.)

The results of the first RAE, initially referred to as a Research Selectivity Exercise, were issued in 1986. To understand the motives underlying this initiative, it is necessary to look back over the previous decade. Through the late 1960s and early 1970s, UK higher education had expanded rapidly, a period of new universities,
massive building programmes and expanding student numbers. However, in the middle 1970s, this “golden age” came to an end, a victim of the international oil crisis and consequent economic turmoil. In response, some observers began openly to question whether the country could afford to expand higher education with teaching and research undertaken across institutions and by individuals without reference to performance-based criteria. Whilst not fully articulated at this time, the implication was that mechanisms would be needed to achieve some degree of selectivity. At the end of the decade, in 1979, the Conservative Government lead by Margaret Thatcher was elected, followed soon after, in 1981, by major reductions in funding for higher education. The shock to the system was huge. Significantly in terms of the future evaluation of research, these reductions were applied selectively, the worst hit being Salford with a 43% cut in recurrent grant. Principles of equality between institutions and similar treatment for all were rapidly being overtaken by views that universities were not equal in their achievement and range of activity, and that such differences should be reflected in resource allocation.

The Thatcher Government launched a number of key reforms in higher education. Underpinning these changes were philosophical doubts about the independence of universities as institutions and practical concerns about cost and how funds were used. These views contributed to fundamental changes in the funding methodology for higher education implemented in 1986, most notably the separate identification of funding for teaching and research in the block grant allocated to universities. Linked with these new arrangements, the allocation of resources for research would be calculated in part on the basis of quality judgements. Thus, the first Research Selectivity Exercise took place in 1985-86 and influenced grants to universities with effect from 1986-87.

Examining the background to the Research Selectivity Exercise, it is possible to identify a range of motives for the evaluating of research:

- For the University Grants Committee (UGC), a methodology was required that would facilitate the allocation of funds to universities, at least in part, on a selective basis. The UGC, and its successor bodies, the Universities Funding Council and now the Funding Councils for England, Scotland, Wales and Northern Ireland, have always maintained the public stance that the RAE is first and foremost a device in order to allocate funds for research within the block grant to universities taking into account achievement and performance. This view reflects an idea that high levels of achievement should be rewarded and that weaker performance, if not explicitly penalised, should not be rewarded to the same level.

- For Government, a methodology was required which provided some higher level of accountability in the use of funds. Ministers complained about “the black hole” of research funding; why did it go to one institution and not another, and what did universities do with it once they received it? Moreover, what were the incentives to enhance output of research? For Government, universities were not alike in their activities and in their levels of achievement, and this should also be reflected in funding levels.
Interestingly, for some senior Vice-Chancellors involved in national policy debates and UGC decision making, there were other, much less public, motives. One former Vice-Chancellor commented that:

“There were a few of us concerned about the future of British science. Funds could not be shared around equally. If we were to maintain our international status in expensive subjects like Physics and Engineering, funds had to be diverted from arts and social sciences to science and engineering, and to the best universities. Research selectivity meant we could move in this direction.”

Here, a sense of hidden strategy, of prioritisation and of re-shaping the higher education system begins to emerge. Significantly, the evidence suggests that such ideas were more developed within institutions rather than at national level. This reflected both a philosophical commitment towards greater concentration and a desire to fund the best activities; it also reflected a degree of opportunism on behalf of themselves and their institutions by some of those involved.

Using the Research Selectivity Exercise of 1986 as the starting point, therefore, it is possible to identify a number of motives for the evaluation of research:

(i) accountability in the use of public funds;
(ii) securing better value for money in public research expenditure;
(iii) rewarding good performance;
(iv) incentivising improved performance;
(v) a device to foster the concentration of research in a smaller number of institutions, thereby creating enhanced critical mass in certain institutions.

It can be argued that the true impact of the Research Selectivity Exercise was not the exercise itself - indeed, many academic staff were unaware that it had ever taken place and certainly did not envisage the long-term significance of the process – but was felt after the results were published in the Times Higher Education Supplement. Suddenly, universities could compare their results; academic staff belonged to departments titled as “outstanding”, “above average”, “average” or “below average”. Such publicity prompted vigorous debate and argument about methodology and the need to justify such ratings. As a result, a further exercise, the first “Research Assessment Exercise” was held in 1989, this time with an assessment scale of 1-5 and a formal set of units of assessment. With the abolition of the binary divide and the award of university title to the former polytechnics, it was decided subsequently that access to research funding should be granted to all institutions. Thus, a further RAE was undertaken in 1992, the first to include the former polytechnics. Not surprisingly, the former polytechnics, without any historical basis of funding for research, performed relatively poorly and pressure built up quickly for a new exercise, which occurred in 1996. The fifth exercise took place in 2001 and the next exercise will take place in 2008 with significant differences from its predecessors.

As the RAE evolved after 1986, further motives for the evaluation of research began to emerge.

• For the Funding Councils, the key motivation has always been the provision of, as far as possible, objective data necessary to inform the allocation of resources in order to support research of high quality in institutions. The Funding Councils have always denied that this represents a policy of concentrating research in a
smaller number of institutions; if that occurs, it is a by-product of the decision to fund high quality research. However, as time went by, the Funding Councils became informal advocates of evaluation for other reasons. UK research performance is very strong by international standards. Total research output, measured by citations, is second only to the USA and output per unit of expenditure is higher than the USA; in many subject areas, the UK can genuinely see itself as the world-leader. Whilst no clear causality can be proved, there is a belief shared by many policy makers that evaluation has led to enhanced performance. This enhancement is associated with benefits of concentration, of competition between institutions and of rewarding high levels of achievement. Funding Council officials are also conscious of another effect of research evaluation, namely the improved internal management of research within institutions. The RAE has had a deep impact on research administration within universities. Senior officers with managerial and leadership responsibilities for research; Research Committees responsible for Research Strategies; close monitoring of research output from academic staff; staff development for research; better management of research careers and research contracts; and schemes to promote research, are all familiar in UK universities, and have assumed key importance in the main research-led universities. There is no longer any hiding place for under-achievement in research. Fully aware of such consequences from evaluations, the Funding Councils have welcomed, often informally, the better management of research which has occurred.

• For Government, as the RAE developed, it began to fit in increasingly with ideas of “new public management” which emerged in the 1980s and 1990s. The research judgements were necessary for a transparent, formulaic allocation of funds linked to quality and accountability; ratings were in line with the perceived need for performance indicators; the RAE promoted competition between institutions which was philosophically welcomed; and the fact that the publication of results and methodology enhanced accountability in the use of funds at both national and institutional levels was a further benefit. As such, the RAE sat very happily within the philosophy and approach adopted by successive Governments towards the end of the twentieth century. The relative success of UK research was also welcomed. Only more recently have long-expressed concerns about cost, not least within institutions; about “game playing”; about problems in the assessment of interdisciplinary research; about the RAE as an end in itself; about the alleged bias towards “pure” rather than “applied” research; and about other technical issues, begun to shed doubt over the value of the RAE, prompting the Chancellor of the Exchequer in February 2006 to announce that the 2008 Exercise would be the last and to initiate an immediate review of alternative arrangements for the future. The development of research ratings also underpinned Government philosophy in other ways, especially the encouragement of institutional diversity. The assumption here is that, by showing some universities to be better in research, others will be encouraged to focus on different strengths, possibly in teaching or in widening access. Similar arguments could also apply within institutions, between departments and even within departments. The ratings also fitted with the developing assessment culture, as illustrated elsewhere by ratings of schools, hospitals and local authorities.

• Of particular interest has been the way the motivation of institutions has also evolved. Especially within research-led universities, the external evaluation of research through the RAE has been used as a driver for strategic change and
Restructuring. Relatively weak RAE scores, with consequent adverse effects on income (both directly from the Funding Councils and indirectly from other sources), became the rationale for change which might or might not have happened otherwise. For some universities, RAE became a useful target on which necessary change could be blamed. A Deputy Vice-Chancellor of a strongly research-led university summed it up as follows:

“The RAE enabled us to change things. We wanted to merge the Departments of X and Y, but needed a reason. The RAE gave us that reason.”

Of course, things could work the other way. Another Deputy Vice-Chancellor commented as follows:

“Getting a 5 was the worst possible thing. We needed to restructure in A but a good RAE score made it more difficult for us to justify change.”

In this way, RAE became a tool for strong managerialism within institutions, both directly in terms of RAE co-ordination and more broadly in terms of academic restructuring. Nor was it just the RAE results themselves; a poor performance revealed in RAE preparations could be equally significant.

Here, therefore, some further motives for the evaluation of research begin to reveal themselves:

(i) improved institutional management, especially in research;
(ii) improvements in the overall level of research, both in quantity and quality;
(iii) encouraging competition between institutions, departments and individuals through publication of ratings;
(iv) an assessment culture;
(v) underpinning of restructuring with institutions.

Overall, it is clear that the motives for the evaluation of research are complex. They may vary between different stakeholders and they may vary over time. This is not unique to the UK and to the RAE. There are important issues here which impact upon the evaluation of research much more generally. Once the motivation is understood more clearly, it becomes easier to assess the methodology. It is also important to recognise that the evaluation of research is, in reality, nothing new or unfamiliar for academic staff. Most countries have some form of competition for the allocation of research funds, and such competition often involves some degree of evaluation and peer review; within institutions, some degree of evaluation of outputs, both in quality and quantity, has always underpinned staff progression and rewards. This leads naturally to issues of methodology as well as motivation.

**Methods**

The need to evaluate research is therefore familiar within higher education. However, methodologies are often highly contentious. The RAE is also especially interesting in that it effectively combines the evaluation of individual staff outputs with an assessment by organisational unit (unit of assessment), leading through the first five exercises to a single rating or score which then provides a weighting for resource allocation and in the next RAE to a research profile by unit of assessment.
RAE methodology developed rapidly in the early years. For 1986, universities were asked to provide information on research income and graduate students, and to provide a sample of research outputs, normally publications, which were seen as representative of the subject area in that institution. Judgements were made by UGC committees and ratings were issued by subject area as “excellent”, “above average”, “average” and “below average”. No formal assessment criteria were used. The exercise was widely criticised for lacking transparency and objectivity; it was widely assumed that many judgements were based on reputation rather than hard evidence. As a result, the exercise lacked credibility within the higher education community.

The 1989 RAE looked very different and was the first to take on the format familiar in subsequent years. Standard forms were introduced for the collection of data by units of assessment, including information on individual members of staff, research students and research income; text based forms were introduced to collect information on research strategy, other measures of esteem and any other relevant factors. For each unit of assessment, universities were asked to identify staff “active” in research and for each member of these staff to cite up to four items of public output; by implication, academic staff not so identified were “not active”. Already, in 1989, some universities appreciated the need for careful planning and presentation of data and information. Central vetting of unit returns, “tactical” decisions about who to include and how to return staff became important and an emphasis on good presentation all began to emerge. Universities will always find ways to exploit a system for maximum benefit and some had already understood the significance of the RAE, both for Government grant but also for wider research esteem. Unit of assessment returns were submitted in paper format before consideration by panels set up specifically to assess the research. In this way, the RAE maintained the approach of “peer review” familiar in other areas of academic life. Results were given on a scale of 1-5, with 5 representing highest levels of international research. For the 1992 Exercise, the methodology changed very little. One innovation was the option to indicate whether research was “applied” in which case the research was separately assessed; but this device proved divisive (there were charges of “second class” research) and was dropped in 1996. Another was the introduction of letter grades, in addition to the number grade, in order to indicate the proportion of eligible staff who were returned as research active. Interestingly, in 1992, information was collected about the total number of outputs in addition to the four named items of output. This prompted particular criticism, with allegations of paper counting, and was dropped as a requirement in 1996. For 1996, at on level, the methodology remained unchanged; in other words, peer review based on a range of information collected. However, the 1996 Exercise marked a new level of professionalisation in the conduct of the assessment. Panels were appointed based on nominations from within the sector in order to increase confidence in the judgements; panels issued clear criteria for reaching their judgements; information was collected electronically as well as on paper; the marking scale was refined to include 3a and 3b and 5*, thereby forming in effect a seven point scale. For the first time, formal arrangements for the audit of university returns were introduced, and uncovered several examples of erroneous returns. In 1996, concern also began to be expressed about the assessment load on panels required to assess a huge quantity of work in a short period.

The RAE of 2001 maintained the same basic methodology as in previous years, with peer review undertaken by panels nomination from within the academic community. Data collection was electronic, with seven sections in each unit of assessment return;
a summary of staffing (RAO); details of research-active staff (RAI); up to four items of output per member of research-active staff (RA2); information on research students (RA3); information on research income by source (RA4); a statement of research strategy and environment (RA5); and a statement of particular circumstances and other distinctions (RA6). Thus, the methodology comprised the collection of information, both input and output, and a peer review judgement reaching a rating by unit of assessment. Each panel set out clear methods of working and criteria, both of which were followed scrupulously (encouraged by a deep fear within the Funding Councils of potential legal challenge. As in 1996, panels drew upon international assessors to confirm the very top judgements of 5* equating to international excellence; some panels also draw upon “users” of research for further input. Efforts were made to ensure consistency across the 67 units of assessment by establishing overarching panels to co-ordinate assessments in a broad subject area. Bearing in mind the “transfer market” of leading researchers moving between institutions, universities were able to claim credit for staff who had recently moved. In total, over 180,000 separate items of research output were assessed.

The 2001 exercise was a very professionally run operation, within the Funding Council and within most universities. By this time, the “rules of the game” had become familiar. A genuine increase in the quantity and quality of research activity combined with a broad understanding of RAE “strategy” (who to select, how to target submissions) resulted in a significant improvement of grades. For the Funding Councils and the Government, this was a problem, because resources were not available to reward the improved performance. Concerns continued to be expressed about the cost of the exercise, especially in terms of staff time; there was a sense that the RAE had become an end in itself, almost more important than the research in question. Continuing concerns about the treatment of applied research; about equal opportunities (there is some evidence that the assessment disadvantages female staff); about alleged “game playing”; and about particular distortions in some subject areas, prompted a major review of the RAE undertaken by Sir Gareth Roberts. The Roberts Review proposed a two stage exercise: in essence, an internal assessment followed by a detailed assessment for some institutions. The aim was to reduce the effort expended by some universities to little benefit. However, this proposal was shelved mainly for fear of creating an explicit two-level higher education system (rather than the implicit multi-tier system which is the actual outcome of the RAE). For the next RAE, basic principles of peer review have been maintained and a similar range of information will be collected. However, a key change for the 2008 Exercise will the use of research profiles showing the proportion of staff by unit of assessment rated at 1*, 2*, 3* or 4* rather than the single grading as in previous exercise.

This paper is not intended to provide a comprehensive account of the development of the RAE or to assess in detail its impact on UK higher education. Rather, the intention here is to use the experience of the RAE to increase the understanding of issues and methods relating to the evaluation of research. An important starting point is to identify some of the key principles that have emerged:

- **Transparency:** After the first Research Selectivity Exercise, which was patently not transparent, the RAE has demonstrated high levels of transparency, with published criteria, nominated panels and open working methods; in 2001, the whole RAE database (every return from every institution) was made publicly available and feedback was provided on each return to the head of the institution concerned. Clearly, any peer review system assumes a degree of informed
subjectivity on the part of the assessors and cannot be perfectly transparent; this is rather similar to the debate between advocates of quantitative research and proponents of qualitative research. If some degree of subjective judgement is required, perfect transparency is impossible. The RAE has always, deliberately, been based in judgements by panels following debate on the basis of information submitted. Given this requirement, the need for transparency wherever possible has been clearly demonstrated and has generally worked to good effect.

- **Credibility:** In any system of research evaluation, the credibility of the assessments is an essential *sine qua non*. Funding Councils have attempted to meet this requirement in the RAE through the involvement of senior academic researchers including international experts. Whilst some critics of the process suggest that panels are “conservative” in their interpretation of high quality research, the RAE has achieved a high level of confidence in the final ratings. Concern has centred about the basic rationale or about details of operation rather than about the overall ratings reached. Of course, there are exceptions to this and departments which feel themselves at the wrong side of a perceived borderline judgement will always feel a sense of injustice. There are other examples where credibility has been “tested”; an example in 2001 was the veterinary science unit of assessment where every university received the same rating. However, for an exercise of the scale of the RAE, the level of acceptance of the results is remarkable. This reflects the clarity of the process, the willingness of panels to share the methodology with their subject communities and, most important of all, the simple honesty of those participating.

- **Simplicity:** In aiming to evaluate research across all subject areas and across a wide range of institutions, the need for simplicity is paramount. The rules of the exercise need to be easily understood by all concerned. This has not always been the case with the RAE. Rules governing for example, the eligibility for staff for inclusion or the income which may be included may appear relatively simple, but in practice conceal depths of interpretation which have given considerable scope for tactics and game playing. Thus, there are many examples of institutions going to considerable lengths to include or exclude certain individuals, or to return individuals in a particular way. Universities are no different from any other organisation; they will always work within a system to maximise their advantage and they will stretch rules to the limit. In such circumstances, simplicity of operation is often the best strategy for the evaluation of research.

The RAE has also highlighted some key issues which need to be addressed in any evaluation of research:

- **Pure and applied research:** A constant criteria of the RAE has been the alleged failure to attach sufficient value to “applied” research and an alleged bias towards “pure” research. In other terms, this may be described as favouring “Mode 1” research against “Mode 2”. Attempts have been made to counter this argument by involving “users” of research in the assessments and this will be strengthened further in the next exercise, but the criticism remains. On occasion, this debate degenerates into an argument between “old” universities (loosely associated with pure research) and “new” universities (equally loosely associated with applied research). The main point here seems to be that, given the objectives of the RAE, the evaluation is seeking to find “the best” research;
The Evaluation of Research: Motives, Methods and Misunderstandings

Professor John Taylor, CHEMPaS, University of Southampton, UK

this is not the same as the most useable. The use made or impact of research are criteria to be used in deciding what is the best, but they are not the only ones. The key to the debate between pure and applied in the evaluation of research lies in defining the objectives of the exercise from the outset.

- **Comparability across disciplines:** The RAE is especially challenging in that, in theory, it seeks to compare research performance across all disciplines. In most countries, there is some degree of evaluation of research (usually in comparing the record of staff competing for grants or contracts), but the evaluation is usually within a single discipline or within a small range of disciplines (eg within, say, economics, or within, say, the social sciences). The RAE, mainly because it was primarily intended to inform the allocation of resources across the entire range of research activity, aims to have a common standard; thus a Grade 5 in, say, History equates to a Grade 5 in Mechanical Engineering or a Grade 5 in Clinical Medicine. This is incredibly difficult to achieve and to sustain. There are further complications: how do you compare, say, a painting or a musical composition with a paper in an academic journal? There are huge differences in the way research is conducted across disciplines: the lone worker compared with groups of researchers; subjects that require significant external funding compared with others that do not; equipment/laboratory subjects compared with library-based; and subjects where the norm is to publish 5-6 papers each year compared with subjects where the norm is the long monograph prepared over several years. There is no doubt that, in undertaking their ratings, assessment panels have reflected their own subject area culture, whilst still trying to fit in with the overall grades. Moreover, some of these differences exist, of course, within as well as across disciplines. Again, the answer seems to lie in defining the objectives of the exercise from the outset. Whilst the funding available remains an overall sum of money to be allocated, rather than specific sums for particular subject areas, an across the board evaluation of research quality is inevitable.

- **The evaluators:** In any evaluation of research, key responsibility lies with the people making the judgements. They must be senior, respected researchers themselves (to provide credibility); they must reflect a range of institutions and research commitments (to provide acceptance); they must meet various social expectations (by race and gender); and they must combine objectivity with energy and commitment. In general, the RAE panels are seen to have worked very hard; allegations of bias or some other form of malpractice are almost unheard of. The quality of the evaluators must underpin any evaluation process, especially one depending on peer review. With the RAE, the main criticisms have been aimed more at the panels as groups, rather than at individuals; such criticism has been concerned with alleged "group" behaviour, to advance "their" subject area relative to others or to apply particular criteria. Thus, for example, some panels are known to have applied rankings by the perceived quality of journals (rather than by the specific piece of research) and to have applied quantitative as well as qualitative measures (a particular issue, for example, for Engineering panels); less substantiated are the suggestions that some panels, having been seen to have been excessively hard in one RAE, and therefore in practice disadvantaged their discipline, sought to rectify the position in the following RAE by being more "generous". It is very difficult to legislate against such behaviour; steps have been taken to minimise such action by introducing overarching panels intended to achieve greater consistency across broad subject areas.
• **The researchers:** A key issue for the RAE has been the need to identify staff as “research active” in order that their output be included and assessed. This has proved to be highly contentious. At one level, especially in leading research-led institutions, not to be chosen (i.e., to be deemed “not active”) can be a serious blight on career development and may even call into question the future of that individual within the institution. This decision becomes all the more contentious when “tactical” decisions also come into play. Thus, a university aiming for a 5 may choose not to return somebody who might be included if it was going for a 4; the question “do we go for a small 5 or a big 4?” is familiar in RAE discussions. Equally, somebody not returned in one institution might have been selected elsewhere. Of course, other forms of evaluation also include some degree of selection, but it is effectively self-selection (individuals do not apply for funds or promotion and are therefore never considered). The issue with RAE is more that such processes are in most universities very open and public; staff selected or not selected are readily identified. There can be very serious human resource issues arising. Equally, advocates of the RAE point to the fact that, since there is no “hiding place” for weak or inactive staff, the overall level of research activity has increased in both quality and quantity.

A number of other human resource issues have arisen from the RAE, including the need to make special provision for newly appointed or younger staff and for female staff whose research track record may have been interrupted by periods of maternity leave or childcare. A significantly smaller proportion of all female staff has been returned as research-active than male staff. A further concern with the RAE has been the development of a “transfer market” in leading researchers, as universities compete with each other to secure the best researchers. This can be unsettling for many institutions, especially when a university has invested in the career development of a top researcher only for them to move when they are at the peak of activity and, for some people, the sight of sometimes several universities participating in an unseemly competition to secure a “star” researcher is unedifying to say the least. However, there is another side to this argument. The RAE has certainly contributed to the increase in academic salaries for top researchers; it can be argued with some justification that the true “value” of researchers is recognised now in a way that was not the case in the past. The alleged transfer market can certainly be exaggerated in scale; moreover, such movement between institutions happens anyway and also occurs in countries without an RAE. Lastly, the RAE has undoubtedly caused institutions to focus on the career development of researchers, from initial appointment, through staff development programmes to ongoing appraisal and support.

• **Interdisciplinary Research:** The need to provide for research at the boundaries between disciplines is critical. The RAE has always struggled to cope in these areas. Provision has been made for research to be cross-referred between panels and overarching panels are also expected to bear this in mind. However, the requirement that an individual researcher can only be returned in one unit of assessment has always forced difficult decisions to be made. Meanwhile, suspicions have always been voiced that panels work to favour research that is central to their discipline.

• **International Research Ratings:** This is another very contentious area. The definition of “international quality” has been hotly disputed especially in highly specialised subjects like English Law or Celtic Studies. Similarly, the use of
international referees has caused problems of interpretation of the RAE and of comparability. A key lesson for evaluation more generally is to exercise caution in the use of such definitions and descriptions wherever these can be ambiguous or misunderstood.

Overall, therefore, there are several broad themes to emerge from the methodology of the RAE which provide guidelines for the evaluation of research in general:

(i) The need to define the objectives of the evaluation from the outset and to communicate these clearly and effectively to the academic community.
(ii) The need to select appropriate evaluators, to support them through training and guidance, and to monitor their behaviour both as individuals and as groups.
(iii) For institutions, to provide full support for academic staff in their preparation for and response to research evaluation, taking full account of the special needs and circumstances of particular groups.
(iv) The need to be aware of differences in research practice between different subject areas.
(v) The need to provide for interdisciplinary research.

**Misunderstandings**

There is clearly much that can be learned from the RAE that can be applied more generally to the evaluation of research. The RAE has been groundbreaking; the scale of the exercise which, once the 2008 Exercise is completed, will have taken place six times is formidable. Views vary widely as to its relative success or failure, but it is guaranteed to prompt argument and debate. Underlying many of the concerns are some fundamental differences of opinion and misunderstandings.

Some academic staff hold to a view that funding for research should be a right for both institutions and individuals, not something to be completed for and evaluated; this view is closely tied in with ideas about the interaction of teaching and research. These staff hark back to days before the RAE with a rather simplistic nostalgia. However, in the present and foreseeable political climate, it is very unlikely that Government would ever sanction the allocation of public funds for research without some degree of evaluation. There is also growing evidence that the younger generation of academic staff, whose career has developed through the RAE years, broadly accept the ongoing evaluation of research. Even in universities without a long tradition of research, there is also a view that the RAE offers an opportunity for strong researchers to compete. Thus, there are now many examples of strong research groups, often highly selective and focused, working successfully in universities without a strong overall record. More generally, this may be seen as evidence of institutional dynamism and change within the system; it also confirms the avoidance of a “fossilised” allocation of research funds. Therefore, whilst in practice, the allocation of resources arising from the RAE is dominated by a relatively small number of institutions (75% of funding goes to 24 universities) there are opportunities for other groups to excel and to be rewarded accordingly.

As noted earlier, a further criticism of the RAE is that it has become an end in itself. Some would go further, suggesting that it has become an obsession with many universities and staff. A related argument is that it has placed an undue emphasis on research at the expense of education (teaching and learning). Some observers note that, especially in research-led institutions, research has become the key driver for
staff with teaching relegated to a second class activity. There is some substance in this argument, as witnessed by accelerated promotions procedures for leading researchers, the emergence of groups of top researchers who undertake little or no teaching and, conversely, increasing teaching loads undertaken by “non-active” staff, teaching fellows and teaching assistants. However, these are not characteristics unique to the UK. Moreover, it can be argued that the scenario painted is as much to do with the failure of the UK Funding Councils and institutions to recognise, incentivise and reward high quality teaching as it has to do with the “success” of the RAE in raising the profile of research activity.

Another criticism of the RAE relates to cost. It is very difficult to present accurate data for the costs of the RAE. Direct costs can be measured, but it is virtually impossible to estimate the costs of staff input and especially to consider the opportunity costs involved. Most research-led universities undertake exhaustive reviews and preliminary or “mock” exercises before the real thing, all of which consume staff time and effort. One suggestion is that the next RAE in 2008 will cost about £100million (approx 130,000,000 Euros) over the seven years since the last exercise in 2001. However, the real cost is likely to be lower because the figure includes expenditure within universities that would have been incurred normally in the internal management of research. A figure of £100million provokes horror among many working in higher education and provides sustenance to critics of the RAE. Yet, here again, there are misunderstandings. Recalling that the primary purpose of the RAE is to confirm the allocation of research funds, the figure needs to be put in perspective; it is less than 1% of the £10.5billion that will be allocated in research funds over the same seven year period.

Nevertheless, concerns about cost persist and prompt the assertion that “there must be a better way”. This conviction lies behind the Government’s decision that the 2008 RAE will be the last; indeed, at the time of writing, there is some doubt as to whether even that Exercise will proceed. Many of the criticisms of the RAE, as has already been demonstrated, are either unfounded or are based on misunderstanding. Much the same applies to many of the alternatives put forward. One option is that the allocation of funds through the Funding Councils should be tied to other metrics, especially income from other funding sources for research (grants and contracts). The argument is that a quality judgement has already been made in allocating the grant or contract and that a second level quality judgement is redundant. Moreover, the Funding Councils already use this approach in allocating capital funds for research. However, there are counter arguments. Most important, levels of income associated with grants and contracts are not in reality a simple measure of quality. Funds allocated through the Research Councils are strictly peer reviewed and do indeed reflect quality, but they are also heavily influenced in subject areas by Government directed funding; funds allocated as contracts often reflect commercial need rather than quality. There are further issues about relating Funding Council grants to other funding sources: an increase in applications for grants from Research Councils which have relatively high administrative costs and low success rates, with consequent high opportunity costs, would arise; greater volatility in the distribution of funding from year to year and between institutions causing instability in the system may result; and researchers may be driven to follow priority areas for research and deterred from less fashionable or “blue skies” research.

The detailed methodology could also vary. Over the years, many different forms of evaluation have been suggested. A key argument is between advocates of some form of judgement based on supplied information and debate – a peer review system.
– and advocates of a more formulaic approach, commonly based on input metrics (such as student numbers and research income) and output metrics (such as papers published and citations). For some critics of the RAE, the problem is that panels do not have enough information on which to make good judgements and the possibility of departmental visits and reviews has been raised. However, in the UK context, such procedures would be both time-consuming and expensive. Other critics question the role of peer review and favour a more formulaic approach. Much debate has centred upon the use of citations as an alternative. Some research has demonstrated a clear correlation between RAE grades and citations, most notably in the life sciences and medicine and in engineering. However, the use of citations can be seriously questioned in the humanities and social sciences. A recent survey of Economics journals over a 25 year period has also suggested that academic staff who publish in less prestigious journals are more likely to be cited than academics who publish in the highest quality, most rigorously refereed journals; the work also shows that one very high quality paper may be cited less than a cluster of weaker papers. This highlights the problem of citation analysis in very stark terms, especially if the aim of the exercise is to evaluate the quality of research. Citations have a limited value in some subject areas, but even this value is at an aggregate level; analysis at the level of a particular department or individual is almost impossible with any degree of accuracy. Also, as with other possible methods, the widespread use of citations as a tool for evaluation may encourage researchers to favour more easily published research in the most “fashionable” areas.

Attitudes to the use of citations vary widely and point to the potential in the future for a different approach to the RAE, one which uses different methodologies for different subject areas, possibly emphasising peer review in the arts and social sciences and the use of metrics in the physical and life sciences, engineering and medicine. This would require an acceptance that the application of a single set of procedures and guidelines is de jure impracticable (it may already be the case de facto).

Thus, much of the criticism of the RAE is based on misunderstandings, either of the RAE or of its process, or misunderstandings of the alternatives. It is relatively easy to criticise the RAE; it is much less easy to develop alternatives which meet the objectives of the present system. This point has been made in every review of the RAE undertaken, forcing many universities and Funding Councils to conclude that, in the absence of any realistic alternative, the RAE should continue. Criticism of the RAE as a methodology for the evaluation of research also stems from vested interests; any process of evaluation is more likely to be favoured by the “winners” than by the “losers”; any alternative system is more likely to be favoured by those who expect to “win” rather than those who expect to “lose”.

To conclude, this paper has sought to use the experience of the RAE in the UK to highlight a range of different issues relating to the motives and methodologies for the evaluation of research. An underlying message is the need to spell out the motives and rationale for the exercise very clearly from the outset and to recognise changes if and when they occur. Another key message is to recognise that universities are full of clever people who will always manipulate and interpret any exercise of this kind to maximum benefit; this will be true of any successor to the RAE just as much as it is true of the RAE itself. A third key message in undertaking the evaluation of research is to beware of unforeseen consequences. The group of individuals who planned the Research Selectivity Exercise in 1986 had no idea what they would unleash. They had no idea about “league tables”; no idea of the wider consequences for institutions
and for individuals; no idea, for example, that some countries would only fund research students to come to grade 5* departments; and very little idea of the power of market forces in higher education. However, what is also very clear is that evaluation of research will remain in some form, to guide the allocation of scarce resources and as a benchmark for quality used by both Government and by institutions themselves in the management of research activity.