How Should Research be Organised? An Alternative to the UK Research Assessment Exercise

Donald Gillies

Published in: **From Knowledge to Wisdom: Studies in the Thought of Nicholas Maxwell**, Edited by Leemon McHenry, Ontos Verlag, pp. 147-168.

1. Introduction

This paper is a sequel to an earlier paper[1] which criticized the UK Research Assessment Exercise (henceforth RAE), and argued that its likely effect is to make the research output of the UK worse rather than better. The aim of the present paper is to complement this criticism by putting forward a positive suggestion, and so, in the paper, I will outline a way of organising research which I think would produce better results than the RAE.

Perhaps it would be as well to begin by describing the RAE for those who are unfamiliar with it. The RAE was introduced in 1986 by Thatcher, and was continued by Blair. It works like this. At intervals of a few years, RAEs are carried out in all the universities of the UK. The first step is to appoint a committee of assessors in each subject. These assessors are usually academics working in the field in question in the UK. Next most members of each department in a subject have to select a set of pieces of their research. The department then submits all these pieces of research produced by its members to the assessment committee. The members of the committee study this research output, and, on its basis, grade the department on a scale running from very good downwards. The departments which score well on the RAE are provided with research funds. Those which don't score so well are less fortunate. They are provided with much smaller funds for research, and the members of such departments have to spend more time on teaching. Recently there have even been moves in some universities to close altogether departments which perform badly on the RAE.

As the RAE is an institution specific to the UK, it might be thought that a consideration of its merits would be of parochial interest only. However, this is not the case. Other countries, such as, for example, Italy, are considering introducing a RAE along the same lines of the UK. Moreover, as we shall see, the RAE involves methods, notably peer review, which are used in most countries for evaluation of research. My critique of the RAE applies to many of these methods. Finally the alternative to the RAE outlined in this paper is a system which could be adopted in any country.

My suggested alternative to the RAE is designed to avoid the defects of the RAE which were described in my earlier paper. Thus to make the present paper intelligible without the need for reading the earlier one, I will briefly summarise my earlier critique of the RAE in section 2, and indeed add a metaphor: 'throwing away the pink diamonds', which helps to capture the key point of this critique. At the end of this section, I will show that the plan of this paper ties in very well with the philosophy of Nicholas Maxwell.

One qualification should now be made. In this paper I will consider only non-laboratory research. This might seem a rather severe limitation since the popular image of research is of white-coated scientists working in a laboratory. However, non-laboratory research is in fact quite an extensive area. It includes all of the humanities such as history, literary and linguistic studies, philosophy, etc. It also includes a good deal of the social sciences, and in particular most of sociology and economics. Then there are disciplines such as mathematics, theoretical physics and computer science. I have two reasons for limiting myself to non-laboratory research. First of all it is the simpler case to consider, and thus is the more natural starting point for trying to devise a system for organising research effectively. It is simpler because the equipment needed, such as libraries, computers, etc., is automatically provided in any university. Thus the problem is only that of choosing the staff who should be allocated time for research. Laboratory research involves additional, and often very expensive, equipment. So there is the further problem of deciding what pieces of such equipment should be purchased. The second reason why I will not consider laboratory research is that I have never taken part in such research either as an individual or as part of a joint project. By contrast I have carried out a great deal of non-laboratory research which has been mainly in my own subject (history and philosophy of science and mathematics), but has also involved working on interdisciplinary research projects with computer scientists, mathematicians, and economists. My long years of carrying out research (so far 42 in number) have taught me that research is a strange activity which often works in quite counter-intuitive ways. Thus it is highly dangerous for anyone without direct experience to suggest rules for how research should be organised. As I lack experience of laboratory research, I prefer not to discuss it. However, I do think that some of the principles developed in this paper for non-laboratory research could be extended to laboratory research, and I hope that someone who is sympathetic to the approach and has experience of laboratory research will carry out this extension.

2. Critique of the RAE. Throwing away the pink diamonds

I will now give a summary of the arguments against the RAE presented in my earlier paper[1]. The RAE relies on what is known as *peer review*. This means that the value of a researcher's work is judged by a group of researchers working in the same field – the 'peers' of the given researcher. Indeed the RAE in a sense involves a double use of peer review. To be entered for the RAE, a work usually has to be published in an academic journal, and most academic journals use peer review to assess whether submitted papers are worth publishing. Then of course the already published work is submitted to the RAE committee for a further peer review evaluation.

There is, however, a major problem with peer review. A study of history shows that it can in some cases go very wrong. It can happen that the majority of contemporary researchers in a field can judge as worthless a piece of research which is later, with the benefit of historical perspective, seen as constituting a major advance. In my earlier paper[1], I consider in detail three examples of major research advances which were judged by contemporary researchers to be valueless. The first is Frege's introduction of modern mathematical logic, which has become an essential tool for computers. The second is Semmelweis's introduction of antiseptic precautions in hospitals such as washing the hands with antiseptic. This is now routine practice, but Semmelweis's suggestions when he introduced were regarded as absurd by the medical community of the time. The third is Copernicus' heliocentric hypothesis which, when it was introduced, was considered absurd not only by the general public but by most professional astronomers of the time. To make matters worse, what the study of history shows is that peer reviews most often go wrong for the really important research advances. Suppose a researcher makes a small, but competent, advance of a routine kind. Peer reviews in such circumstances will usually be able to give his or her work a reasonable evaluation. When, however, a researcher makes an advance which is later seen as a key innovation and a major breakthrough, peer review may very well judge it to be absurd and of no value.

An analogy will help to explain this crucial defect of peer review. Suppose we have a system to separate flawed diamonds, which have little value, from clear diamonds which are valuable. This system, let us suppose, works very efficiently in eliminated worthless flawed diamonds, but then it turns out to have a crucial defect. As well as eliminating the flawed diamonds, it eliminates the pink diamonds, and pink diamonds have a value a thousand times greater than that of the ordinary clear diamonds. Once our system had been found to have this defect by diamond producers, they would hastily stop using it. My claim is that systems based on peer review, such as the RAE, have exactly the same defect. They are liable to throw away the pink diamonds.

I try in my earlier paper to clarify this point by introducing[2] the distinction between Type I and Type II errors. A research assessment procedure commits a Type I error if it leads to funding being withdrawn from a researcher or research programme which would have obtained excellent results if it had been continued. A research assessment procedure commits a Type II error if it leads to funding being continued for a researcher or research programme which obtains no good results however long it goes on. In terms of our analogy with sorting diamonds, throwing away a pink diamond is committing a Type I error, while retaining a genuinely flawed and hence valueless diamond is committing a Type II error. Now the problem with the RAE is that it concentrates exclusively on eliminating Type II errors. Yet the history of science shows that Type I errors are much more serious than Type II errors. The case of Semmelweis is a very striking example. The fact that his line of research was not recognised and supported by the medical community meant that, for twenty years after his investigation, thousands of patients lost their lives and there was a general crisis of the whole hospital system.

But how is it possible for peer reviews to go so wrong, and to throw away pink diamonds, i.e. to judge as worthless what are later seen as major advances in the subject? At first it may seem paradoxical that this should occur. After all, the peers, who do the reviewing, are all experts in the field and active researchers. Surely they, of all people, should be able to recognise good research when they see it. Despite the apparent strangeness of this situation, the reasons why it occurs can in fact be quite well explained using ideas from the philosophy of science, more specifically using Kuhn's paradigms, and Lakatos' research programmes.

Kuhn argues that scientific research is usually carried on by groups of scientists who all accept, as the basis of their research, a general framework of assumptions, called a *paradigm*. In periods of 'normal science', the correctness of the paradigm is never questioned. Only in occasional revolutionary periods is the paradigm criticized, and a scientific revolution can result in an old paradigm being replaced by a new paradigm. Once the revolutionary period is over, however, normal science resumes but on the basis of the new paradigm rather than an old one. Given this model of scientific development, it follows that most researchers in a period of normal science will regard as absurd any development which contradicts the dominant paradigm. This provides a neat explanation of why the ideas of Frege, Semmelweis, and Copernicus were rejected by most contemporary researchers in the field. Later, however, when a new paradigm has been accepted, the ideas of Frege, etc. seem obvious to the experts in the field and they find it difficult to understand why they were rejected earlier. In terms of peer review, we can say that any idea contradicting the dominant paradigm is very likely to be rejected by peer review, but many such ideas will be seen later as introducing revolutionary advances in the subject.

However, the failure of peer review need not be exclusively associated with scientific revolutions. It can occur in what Kuhn calls 'normal science' as well. To see this, let us suppose that research is being carried out on some problem and that four different research programmes have been proposed to solve it. We can further suppose that all four of the programmes are compatible with the dominant paradigm, so that we are not dealing with revolutionary science. It may be almost impossible to say at the beginning which of the four programmes is going to lead to success. Suppose it turns out to be programme number 3. Let us suppose further (which indeed is often the case) that initially programme 3 attracts many fewer researchers than programmes 1,2 &4. Now it is characteristic of most researchers that they think their own approach to the problem is the correct one, and that other approaches are misguided. If a peer review is conducted by a committee whose researchers are a random sample of those working on the problem, then the majority will be working on programmes 1,2 & 4, and are therefore very likely to give a negative judgement on programme 3. As the result of the recommendation of such a peer review, funding might be withdrawn from programme 3, and the solution of the problem might remain undiscovered for a long time. This example shows not only why peer reviews can give the wrong answer, but also that Type I errors are more serious than Type II errors. Suppose programme 3 is cancelled in order to save money (Type I error), then all the money spent on research in the problem will lead nowhere. It will be a total loss. On the other hand if another programme (5) is also funded, the costs will be a bit higher but a successful result will be obtained. This suggests that funding bodies should make sure that some funding at least is given to every research school and approach, rather than concentrating on the hopeless task of trying to foresee which approach will in the long run prove successful.

So we can sum up as follows. The main defect of the peer review system is that it is likely to throw away pink diamonds. Conversely anyone trying to design a system for organising research should ensure that the probability of throwing away pink diamonds is made as low as possible. Any suggested system for organising research should in my view be subjected to what I call the WFS test. This test consists in taking some leading research achievements from the past, and seeing whether those who carried them out would have fared well under the proposed system. If they would have fared well, the system passes the test. If they would have fared badly, it fails the test and should be altered. In the light of my earlier paper[1], I consider the achievements of Wittgenstein (W), Frege (F), and Semmelweis (S). However, obviously other examples could be chosen. So far I have mentioned Frege and Semmelweis. The problem with Wittgenstein was a different one. Wittgenstein published nothing during the last 17 years of employment at Cambridge. On the current RAE system, he would have been classified as research inactive, and have had his research time removed and would perhaps even have been sacked. Yet during these 17 years, Wittgenstein wrote the *Philosophical Investigations*, which many regard as the best philosophical work of the 20th century. To generalise from this case, we want our research system to allow researchers, if they are so inclined, to spend a long time polishing and re-polishing their works before publication. We know that this strategy can sometimes result in durable masterpieces.

The conclusion to be drawn from these various arguments is that the RAE is likely to encourage working within standard paradigms and mainstream research programmes, and making small contributions. It is likely to discourage new approaches and minority research programmes, and so tends to eliminate major innovations (pink diamonds). It is likely to encourage boring routine research at the expense of interesting, novel and exciting research. This analysis of the RAE ties in very well with the philosophy of Nicholas Maxwell, as I will now explain.

It is a very great pleasure to contribute a paper to this volume in honour of Nick Maxwell since we have been friends during nearly all of those 42 years when the two of us have been carrying out research in London. We have met on numerous occasions and had as many stimulating discussions. What has made these discussions with Nick so valuable for me is that we always seem to agree on many things but not everything. Our two positions have a lot of overlap, but do not quite coincide. There is enough in common to make a discussion possible, with enough divergence to make it stimulating. I have chosen a paper on the RAE for this collection since Nick and I certainly agree in general attitude to the RAE, but yet, as usual, there is a difference as well. However, to discuss this matter more fully, let me revert to the formal academic style of ‘Maxwell’ rather than ‘Nick’.

Maxwell discusses the RAE in the second edition (2007) of *From Knowledge to Wisdom. A Revolution for Science and the Humanities*. He makes encouraging remarks about my own paper on the subject, and goes on to reinforce my argument by giving a further series of examples of scientists and mathematicians whose work was not recognised for many years:

> There are many other cases of people making important scientific or intellectual contributions and receiving no recognition for their work for twenty years or more. Thomas Young's discovery of the wave character of light via his interference experiment was initially dismissed by his peers. Gregor Mendel's discovery of some basic laws of genetics famously had to wait several decades before it received recognition. This was true, too, of Alfred Wegener's theory of continental drift, and John Waterston's contribution to statistical mechanics. Georg Cantor met with opposition when he developed set theory – of profound importance to the whole of mathematics. E. Stückelberg failed to receive recognition for his important contributions to quantum field theory. And Guy Callendar failed to convince when he announced in 1938 that increased emissions of carbon dioxide as a result of human activity was leading to global warming. These cases, I am sure, merely scratch the surface.[3]

Maxwell then goes on to relate the RAE to the question of wisdom-inquiry. Maxwell's general position is that current academic life is dominated by a philosophy of knowledge and that there should be a shift towards a philosophy of wisdom. However, he thinks that the RAE will impede this desirable development, saying[4]: 'It may well be especially difficult for a revolutionary ideas like that of wisdom-inquiry to get a fair hearing in an academic world constrained by the RAE.' Maxwell also gives a striking specific example of a case where the RAE is inhibiting wisdom-inquiry. He writes:

> But how, it may be asked, may the RAE impede acceptance and implementation of wisdom-inquiry? To begin with, as long as knowledge-inquiry intellectual standards are in place, the RAE will make it even more difficult to do wisdom-inquiry research. The point was made to me in a striking way by Dr. Caren Levy, director of the Development Planning Unit at University College London. Her work and research, like those of others in her Unit, is concerned to help the poor tackle their problems of living in Africa and Asia. Here, if anywhere in academe, wisdom-inquiry is being put into practice. But this creates a dilemma. On the one hand, Dr. Levy can publish papers in relevant academic journals, which gain recognition by the RAE but may not lead to anything of value in the real world. On the other hand, reports produced by Levy, dealing with developmental problems in Africa and Asia, widely read by many grappling with these problems, taken up and implemented by the UN and other organizations, and having practical consequences of value in the real world, receive no recognition from the RAE at all, because the relevant reports are not published in academic journals acknowledged by the RAE. In this way, the RAE increases the pressure on academics to produce orthodox, and often useless, knowledge-inquiry work – pressure, I hasten to add, which Levy resists (even if other departments do not).[5]

This criticism of the RAE seems to me entirely correct. Moreover the system proposed here as a replacement of the RAE would, I believe, genuinely help researchers like Dr Caren Levy do their valuable work.

It is clear from this that Maxwell and I are in very broad agreement regarding the RAE, and yet there is still, I think, a difference between our positions – even if this difference is perhaps one of emphasis. The difference is this. My own criticisms of the RAE and suggestions for an alternative are based very strongly on considerations which may not have the same weight for Maxwell. These considerations are a desire for *efficiency* and a wish that research should lead to *wealth-generating* innovations. A system of organising research should, in my view, be as efficient as possible where efficiency is measured by the amount of good quality research produced per dollar, euro, or pound put into financing the system. The RAE reduces efficiency in this sense because it is costly both in money and the amount of time researchers have to devote to it rather than to getting on with their research. Moreover the RAE reduces the amount of good quality research produced. My proposed alternative is designed to be more efficient. Regarding wealth-generating innovations, the most striking of these (as I discuss in my earlier paper[1]) are made possible by the real breakthroughs in research (the pink diamonds). The RAE, however, reduces the probability of such breakthroughs by forcing researchers into routine, incremental, research, and so it reduces the probability of important wealth-generating innovations.

This interest in increasing efficiency and helping to promote wealth-generating innovations puts my thinking more in line with the more standard government approach, since almost all governments profess these goals, even if their policies are sometimes a hindrance rather than a help to achieving them.

At this point, however, it might begin to seem that there is no way of solving the problem of designing an efficient way of organising research. The key problem is that we cannot properly judge the value of a research contribution until about 30 years after it has been made. Contemporary valuations are shown by history to be very misleading. Contributions which in the perspective of history are seen to be major advances are sometimes judged to be valueless by contemporary researchers. Conversely research which seems brilliant at the time is often seen later to be merely a passing fashion which proved to be of no significance in the long run. If, however, we can only evaluate research properly after 30 years, how can we decide *now* what research to fund. The problem does indeed look insoluble.

But the problem is not really insoluble. All that is needed is a new approach, which I will now explain in outline and then elaborate in the rest of the paper. What we need to do is to shift our focus away from research to another activity which nearly always accompanies research, namely teaching. So far we have considered the effects of the RAE on the research output of the UK, but what about its effect on teaching in the UK universities? In the next section (3) I will argue that the RAE as well as damaging the UK's research output also damages the teaching in UK universities. It is just as bad for teaching as it is for research. However, this result suggests the following idea. The RAE was designed to improve research, but ended up damaging both research and teaching. Suppose now we shift from an attempt to improve research to an attempt to improve teaching. As the results of teaching are more straightforward to assess, this might be an easier problem to deal with. Moreover, if we solve it, it could be that the new system designed to improve teaching might as a spin-off (so to speak) improve research as well. It could be that the connections between teaching and research are such that improving one will result in an improvement in the other, just as damaging one results in damage to the other. I believe that this is really the case, and will show that a system proposed to improve teaching will, as an indirect consequence, improve research as well. This will be done in section 4. So, in a nutshell, the system I propose is designed to improve teaching, and it will be shown that the effect of this improvement will be to improve research as well.

3. Why the RAE makes teaching worse

There are three separate reasons why the RAE has a bad effect on teaching in the universities, and I will deal with these in turn. The first, and perhaps most obvious reason is concerned with the reward and hence incentive structure introduced by the RAE. Departments get more money if they do well on the RAE, and have their budgets cut if they do badly. However, whether the department's teaching is good or bad has little, if any, effect on its income. In these circumstances, economic rationality dictates that departments should concentrate their efforts on doing well at the RAE, but not bother so much about teaching. Of course humans are not entirely motivated by economic rationality. Many academics feel they have a professional duty to teach the students well, and this sense of duty may counteract the dictates of economic rationality. Still economic rationality is bound to have some effect, and so the incentive structure introduced by the government in the shape of the RAE is bound to have a negative effect on teaching.

Against this argument, it might be pointed out that the government has also introduced an assessment of the quality of teaching in departments, the so-called QAA, and this will ensure, so it could be claimed, that there is no decline in the quality of teaching. The problem here, however, is that there are rewards for doing well on the RAE and penalties for doing badly, but no such rewards and penalties exist for the QAA. Hence once again economic rationality dictates not paying much attention to the results of the QAA.

But why are there are no rewards and penalties for performance on the QAA? Could they not be introduced? A little reflection, however, shows that the two cases are not symmetric. Suppose a department does well on the RAE. It gets a larger budget and this translates into the staff getting more research time and having to do less teaching. Given the present structure of universities, this is interpreted as a reward. Similarly doing badly on the RAE, and hence having less research time is interpreted as a punishment. But now suppose we wanted to introduce similar rewards and punishments for performance on the QAA. How could it be done? If a department did well on the QAA, this should, if the two cases were really parallel, result in the staff having more teaching time. However, unfortunately, this would be interpreted as a punishment, and so departments would endeavour to do badly on the QAA in order to escape this punishment. But could we then give more teaching to those departments which did badly on the QAA, and less teaching to those which did well. This would give the correct incentive for departments to do well on the QAA, but the net result would be that teaching would be done more and more by those departments which were bad at teaching. This is hardly desirable. The net result is that a department's performance on the QAA has little effect, and the QAA is largely just another time-wasting bureaucratic exercise.

Let me now go on to the second reason why the RAE has a negative effect on teaching. This is connected with a curiously out-dated feature of the RAE. The RAE assesses departments, thereby presupposing that departments are the units of research. Now 30 or 40 years ago, that was largely the case. Research schools were indeed principally located in specific departments. However, during the last 20 or so years, this has been completely changed by the development of globalisation. Everyone knows that globalisation has transformed the world economy, and it has similarly transformed research. Typically nowadays research groups, far from being located in single departments, are scattered throughout the world. Their members communicate on a day to day basis by email, and meet regularly at international conferences.

I can illustrate this change by own experiences. When I started research as a graduate student working for a PhD, I joined the Department of Philosophy, Logic and Scientific Method at the London School of Economics in 1966. The head of department was then Professor Sir Karl Popper, and my supervisor was Imre Lakatos. At the time this department was indeed the centre of a very distinctive research school in history and philosophy of science and mathematics. In the last 20 years, I have continued to do research in the history and philosophy of mathematics, but I have never had a colleague in my department who was researching in this particular area. Did this mean I was isolated and had no one with whom to discuss the problems of the field? On the contrary, I have had many more discussion partners in the last 20 years than I did in 1966. The only difference is that, far from being in the same department, they are located all over the world. Of course these days that does not prevent regular discussions by email, and regular meetings in diverse places. This is well-illustrated by listing some of the collections of papers produced by this lively and stimulating group of researchers. I edited one such collection: *Revolutions in Mathematics*, which was published by Oxford University Press in 1992.[6] There were 12 authors – no two of whom were in the same department. By location 1 was from China, 3 from Germany, 1 from Italy, 3 from the UK, and 4 from the USA. However, 1 of those located in Germany was an Italian national, as was one of those located in the USA. As can be seen, we have here a truly multi-national research group. Subsequent collections of papers in this field tell the same story. One published by Kluwer in 2000 was entitled: *The Growth of Mathematical Knowledge*, and was edited jointly by Emily Grosholz of the Pennsylvania State University in the USA, and Herbert Breger of the University of Hannover, Germany[7]. It was the revised proceedings of conferences held in Pennsylvania State University in 1995 & 1996. Next in the series is: *Mathematical Reasoning and Heuristics* published by King's College Publications in 2005. This was edited by Carlo Cellucci of Rome University and myself[8]. It contains the revised proceedings of a conference held in Rome University in 2004. I mention this example from my own experience because I believe that it is typical. The research group based on a single department or university has largely disappeared to be replaced by multi-national research groups. This parallels the increasing transformation of national companies into multi-national companies. From my own experience I would say that this new form of research organisation is much superior to the old. It allows a much wider range of contacts and discussion partners than did the old system, and this is very important in specialised fields. It also leads to much better human relations within the group, and many fewer quarrels. If a number of researchers, in the same field but holding different opinions, see each other every day in the same department, the outbreak of quarrels in more or less inevitable. (Such quarrels were a striking feature of the philosophy department at the London School of Economics in the late 1960s.) Moreover the situation is made worse by the fact that members of the same department are often competing against each other for promotion. Discussion by email and occasional meetings cools the situation, and makes such quarrels less likely. Moreover the members of a multi-national research group are not competing against each other for local promotions. Indeed they can help each other to obtain such promotions. If the group as a whole succeeds internationally, its members are more likely to succeed in their own countries. All this leads to a more pleasant and constructive atmosphere within the research group.

Whoever designed the RAE was, however, obviously not aware of these developments, and, as a result, the RAE introduces incentives which have a negative effect. Although university departments are ceasing to be centres for research groups, they remain centres for teaching. Students join a particular department and take most of their courses within that department. If a department is offering degrees in a particular subject, then it is in the students' best interest for it to make appointments in every branch of the subject. Each branch will then be taught by a specialist in that particular field, who will know more about it, and be able to present the results better. This appointments strategy which is best for teaching, is, however, undermined by the RAE. If a department wants to do well in the RAE, it has to present itself as being an international leader in some particular branch of research. To do so, the best appointments strategy is to appoint a large number of staff in one particular branch so that it can then claim to be internationally known for that speciality. So, for example, a mathematics department might appoint researchers in a particular branch of mathematics, say category theory, as half its staff. It can then claim to be a world leader in research in category theory. The problem here is that students taking a mathematics degree in that department do not want half their courses to be in category theory. They want a broad coverage of the subject. This means that either the syllabus of the degree offered will be highly distorted, or half the courses will be taught by people who are not specialists in that field. Either way, teaching is bound to suffer. Many departments in the UK have followed just such an appointments policy in order to shine in the RAE, and this must have had a negative effect on teaching. The irony of the situation is that researchers in a particular speciality are being collected into a single department just at the time when the emergence of globalisation and multi-national research groups makes this quite unnecessary and indeed undesirable. It is interesting to note here how damage to research is going hand in hand with damage to teaching. The RAE is inhibiting the development of the superior multi-national research group, while, at the same time, encouraging an appointments policy which is worse for teaching students.

I now pass to the third reason why the RAE is damaging teaching in UK universities. I will explain this by a fictional example which is in fact based on a real life case. It will be clear to the reader why I do not want to give the details of the actual characters involved. My fictional characters are a Ms A and a Mr B. Ms A is a good researcher but a bad teacher. She has a brilliant facility for generating new ideas and lines of research.

She can quickly turn out papers which are admired by her peers and published in the top journals. Unfortunately, however, Ms A cannot get her ideas across very well to a student audience. She is rather shy and tongue-tied, and cannot understand the difficulty which some students have in grasping points which to her are obvious. Mr B is exactly the opposite. He is highly studious, and knows his subject well. Unfortunately though, he just doesn't seem to get many new ideas when it comes to research. Moreover while Ms A can dash off a research paper in a couple of days, for Mr B writing is a slow and painful process and he may take months to complete a short article. On the other hand, once in front of a class of students, Mr B is in his element. He speaks well with a great command of rhetoric. He is a charismatic figure loved by the students, and expounds even the hardest points so clearly that every one can understand them. In this fictional example, based on reality, common sense obviously dictates that Ms A should do more research and less teaching, while Mr B should do less research and more teaching. Before the RAE, such an arrangement would without doubt have been made informally. I will now show, however, that a desire to perform well on the RAE could lead to exactly the opposite allocation of research time.

The rules of the RAE have varied over time, but in most RAEs, including the present 2008 one, there has been an upper limit to the number of items (papers or books) which an individual member of staff can submit. In the current RAE it is 4. Moreover the more members of staff of a department who can submit to the RAE, the better the rating of the department is likely to be. Now suppose, to continue our fictional narrative, that there is a five year period from the last RAE to the next one, and that a year and a half has elapsed. The prodigious Ms A has already completed 4 brilliant papers, while Mr B the poor researcher and slow worker hasn't even managed to complete one. The head of department now considers what must be done to get the best rating for the department at the next RAE. Ms A has already completed all that needs to be done, and to a very high standard. Thus there is no point in giving her more research time. Mr B is a slow coach, but, if he is given a lot of extra research time, he might just manage to get the necessary 4 papers completed. Thus the rational policy for doing well on the RAE is to cut Ms A's research time and give her more teaching, while allocating more research time and less teaching to Mr B. Thus the requirements of the RAE lead to a strategy which makes the teaching much worse for the students, and, at the same time, reduces the quality of the department's research output.

Having shown that the RAE reduces the quality of teaching, I will argue in the next section that measures to improve teaching will have a positive effect on research as well.

4. Why rewarding teaching will improve research

I am dealing in this paper with non-laboratory research. The vast majority of such research is carried out in universities by academics who are not exclusively research workers. In fact these academics generally have 3 rather different activities in their work, namely (1) research, (2) teaching, and (3) administration and management (which I will abbreviate to admin). There are typically 4 grades in the academic hierarchy. In the UK, academics start at the lecturer grade, and can then, if they are fortunate, obtain promotions to senior lecturer, reader, and finally professor. Most universities internationally have four grades of this type, though they often have different names. As regards obtaining promotion, however, there is a very striking difference between the 3 academic activities just listed. Promotions can be obtained for either research or admin, but rarely, if ever, for teaching. An academic who concentrates on teaching might, if he or she is lucky obtain a promotion from lecturer to senior lecturer, but, generally speaking, that is as high as he or she can hope for. To get to a readership or professorship, success in research or admin is what counts. Typically someone might be at the reader grade, and then after taking on an important admin job, such as head of department, would obtain a professorship. Of course professorships are sometimes obtained just by success in research with little admin activity. However, admin is usually a better bet than research for climbing the ladder. The highest grade a research specialist is likely to obtain is professor, while a specialist in admin can go to higher and better paid jobs such as dean, pro-vice-chancellor, or even head of the whole university.

I remarked earlier that teaching is at present of low status in UK universities, and having to do more teaching is usually regarded as a punishment. This is not, I claim, because teaching is at university level is an intrinsically unpleasant or unrewarding activity. Quite the contrary is the case. Teaching bright students who give a stimulating feedback, preparing new courses covering the latest discoveries in the field, etc., all these can be very enjoyable and intellectually demanding. Indeed, generally speaking, teaching is a much more interesting activity than admin. Still ambitious and talented academics prefer to do research or admin, and this is for the obvious reason that work in research or admin can enable a person to climb the status hierarchy, while work in teaching cannot.

How can this situation be put right? There is an obvious way. The criteria for promotion must be altered so that it is just as easy, or perhaps even easier, to climb the ladder by working as a teacher as by working at research or admin. Actually it is much easier to assess a person’s performance as a teacher than to assess someone’s performance as a researcher. The quantity of teaching carried out in terms of number of hours and number of students is immediate. As usual, an assessment of quality is more problematic, but not nearly as problematic as is the case with research. One can use student feedback, and exam results. Then there are more subtle and important criteria, such as those of introducing new teaching methods, up-dating courses to contain the latest results in the field, and so on. Note, however, that I am not arguing for teaching only posts. In fact I will argue against such posts in a moment. My argument is that an academic’s teaching activities should count just as much towards promotion as his or her research and admin activities. If this were to happen, the status of teaching would rise, and the quality of teaching would improve. I will next show that this would also improve the quality of research.

I will now give an outline of the system I propose. As we have seen, academics have 3 activities between which they divide their time, namely (1) research, (2) teaching, and (3) admin. The problem is how research time should be allocated. The RAE attempts to solve this problem along the following lines. An assessment is made of how good an academic department is at research and those departments which are better are allocated more research time. Then, following assessments within the department itself, this is translated into more or less research time for individuals. There are two fundamental difficulties with this approach. First of all carrying out these assessments is very costly, and secondly their results are very dubious. It is perfectly possible that someone who is really a brilliant researcher could end up being allocated little or no research time, thereby throwing away a pink diamond. Instead of this approach, I therefore propose that academics themselves should decide whether they want to do more research, more teaching, or more admin. This I call the principle of *self-selection*. It will be seen at once that this principle solves both the main problems besetting the RAE. First of all it becomes unnecessary to carry out a complicated assessment of all researchers, and so there are enormous cost savings. Secondly, the risk of committing a type I error (throwing away a pink diamond) is reduced almost to zero. It is not possible to recognise the pink diamonds of research immediately, but they all have one characteristic in common. They love research and are very keen to do it. Thus, given a principle of self-selection, they would all go for the research option.

However, at this point, many readers will perhaps smile and conclude that I am proposing a purely utopian scheme which has an obvious flaw. Surely, the objection will be made, if a principle of self-selection is adopted, everyone will opt for research whether they are pink diamonds or plain incompetents, and so it will be impossible to get any teaching or admin done. This objection has an apparent force because there is a certain hypocrisy in academic circles. I have yet to meet an academic who did not claim that what he or she really loved best was research. However, observing the behaviour of those who make such professions, one has to conclude that they are often false. Academics typically start with great enthusiasm for research, but, after a number of years working at research, they often become rather bored with it. They may have run out of ideas. They may have come to realise that their youthful hopes of becoming the next Einstein were an illusion, while the reality is that there are quite a number of younger researchers doing better than they are. In these circumstances the sensible move is into administration and management where a tempting career ladder stretches before them. Indeed many who do switch from research to admin may have carried out quite a lot of brilliant research, but do not think they have the capability of continuing. This is often the case with mathematicians who characteristically make their best contributions when young. Several eminent mathematicians who have proved deep theorems in their youth switch in middle life into administration and management and are often successful at that as well. In so doing they are following in the footsteps of the great Sir Isaac Newton. His masterpiece *Principia Mathematica* was published in July 1687 when its author was 44. However, Newton did not spend the rest of his life carrying our research in mathematics and physics, but rather switched into a new career in administration and management, not in Cambridge University but at the Royal Mint in London. He became Warden of the Royal Mint in 1696, and was promoted to Master in 1699. Apparently he ran the Royal Mint very well.

What I am saying is that many academics would be quite happy to switch their activity away from research provided they do not lose any status thereby and indeed have instead the chance of climbing a career ladder. Given the present set-up, such academics will choose the admin path and shun teaching which gets nowhere. However, if a switch into teaching could lead to a successful career, many academics would be happy to follow that path, and indeed many might much prefer it to the admin path.

These considerations lead to the following outline of my suggested system. A young academic is given a first appointment with a fairly generous allowance of research time. Later on, however, he or she can decide whether to continue with that amount of research time or to switch to doing more teaching or more admin or more of both. The incentive for doing less research and more teaching or admin would be that it would make it easier to obtain promotion and extra money. Obviously the easier it was to gain promotion and extra money by doing say teaching, the more academics would choose the teaching option. Hence the difficulty of getting promotion in the various kinds of academic activity could be adjusted empirically, so that, overall, the required amount of research, teaching and admin was carried out. To put it another way, academics would be tempted into doing more teaching and less research by the possibility of climbing the ladder more easily in this way, rather than, as in the present system, prevented from doing research because their department obtained a low rating on the RAE. However, though some academics would do less teaching and more research and others more teaching and less research, I would argue that all academics should do some of both. I strongly oppose the idea of teaching-only posts for the simple reason that they would result in a decline in the quality of teaching. Suppose someone is appointed to a teaching-only post. In the first few years he or she might be excellent, but, after twenty years, the subject would have moved on. With no research time to study the new developments our 'teacher-only' would inevitably have become out of date, and his or her teaching would suffer. Besides this, teaching in the final year of an undergraduate degree and at master's level involves setting students research projects, and a teacher with no links to research would be unable to do this satisfactorily. The situation would be made even worse by the creation of teaching-only universities.

So, if we want to improve teaching rather than make it worse, it is necessary that every university teacher should have an allocation of research time. This time need not necessarily be used for writing new papers and books (what could be called 'active research'). It could be used for studying the latest developments in the field, attending research seminars etc (what could be called 'study research'). Obviously carrying out active research requires also carrying out study research, but it is possible to do only the latter. Conversely I believe that all researchers should do some teaching. It is very helpful to any researcher trying to formulate new ideas to try to expound them to a student audience. Incomprehension by the students can be an incentive to improve the clarity of the formulation, while often students make critical comments and useful suggestions which lead to an improvement of the content of the research. Thus all academics should do some research and some teaching, but it is perfectly legitimate and indeed desirable that some should do more research and less teaching while other should do more teaching and less research. Remember the story of Ms A and Mr B given in the previous section. We have already indicated how promotions could be based on the quantity and quality of teaching. Promotions on the basis of admin could remain as they are now, while promotions for those specialising in research would be based on peer review carried out in something like the following manner. Researchers who wanted promotion on the basis of a number of years' work since their last promotion would prepare an account of the research they had done with a list of books and papers published, and then submit this together with a selection of say the 3 or 4 of their publications which they consider to be the best. This submission would then be sent round to a number (say 3) of researchers in the same field for evaluation.

Now it may seem rather inconsistent on my part to re-introduce peer review at this point, having criticized it earlier. However, there is no real inconsistency here. The problem with peer review in the RAE was that it might result in throwing away the pink diamonds. However, there is no such consequence here. A 'pink diamond' might fail to gain promotion on the present system, but there would be no question of him or her being 'thrown away', i.e. prevented from doing further research. Moreover pink diamonds are always recognised by their peers after the passage of a number of years. So, in most cases, pink diamonds would obtain promotions after a delay of some years. All this will become clearer if we now apply what I earlier called the WFS test to the proposed system for organising research.

How would Wittgenstein, Frege and Semmelweis have fared in the proposed system? In fact all three would have been allowed to continue their research. Wittgenstein by refusing to publish anything would have failed to obtain promotion and so remained at a low grade. However, this would have suited him very well. Although heir to one of the largest fortunes in Europe, he gave all his money away and liked to live in what can only be described as 'ostentatious poverty'. To have a low-grade and poorly paid academic job would have suited him down to the ground. Frege's fate under the suggested system would be much what it was in real life. He was never recognised by his academic peers during his years at Jena University and in fact was never promoted to the highest grade (Professor Ordinarius). As it took nearly 40 years for his work to become generally recognised, then, in the suggested system, he would have had great difficulties in getting promotion on the basis of peer review, but he would still have been allowed to continue his research. However, Frege is really exceptional from this point of view. While it is not so uncommon for innovators to have to wait some years for their work to be recognised, the number of these years is usually less than twenty or thirty, so that an innovator who, like Frege, produced a great research work at 31 would usually have received enough recognition to climb to the top of the promotion ladder before retiring at 65. Semmelweis died at only 47, but had he lived to be 60, i.e. until 1878, he would have seen his approach become generally accepted by the medical profession, and would, without doubt, have had all the honours, promotions, and acclaim which he failed to obtain because of the shortness of his life. So we can conclude that the suggested system passes the WFS test which is so strikingly failed both by the RAE. As it would also be much cheaper than the RAE, it must surely be judged to be superior.

Notes

1. D. Gillies, “Lessons from the History and Philosophy of Science regarding the Research Assessment Exercise,” in Anthony O’Hear (ed.), *Philosophy of Science* (Cambridge: Cambridge University Press 2007), pp. 37-73.

2. *Ibid*., pp. 60-61.

3. N. Maxwell, *From Knowledge to Wisdom. A Revolution for Science and the Humanities* (Second Edition, London: Pentire Press 2007), p. 317.

4. N. Maxwell, op. cit., p. 318.

5. N. Maxwell, op. cit, p. 317-18.

6. D. Gillies (ed.) *Revolutions in Mathematics* (New York & Oxford: Oxford University Press 1992).

7. E. Grosholz and H. Breger (eds.) *The Growth of Mathematical Knowledge* (Dordrecht, Boston, London: Kluwer 2000).

8. C. Cellucci and D. Gillies (eds.) *Mathematical Reasoning and Heuristics* (London: King’s College Publications 2005).

University College London,
24 June 2008.