

# Measured merit

## Mega-evaluation based on peer review to guide distribution of funds to UK universities

FABRÍCIO MARQUES

*Published in February 2009*

**T**he academic community in the United Kingdom is experiencing a phase of definitions. In late 2008, the results of the sixth Research Assessment Exercise (RAE 2008) were released. This was a major effort to assess the quality of research, in order to determine how US\$2.3 billion a year of public funds will be distributed among British universities from 2009 to 2014. An evaluation of 52.4 thousand academics from 159 higher education institutions was conducted. The findings indicate that 17% of their research studies are at the global leadership level; 37% are in the international excellence category; 33% enjoy international acknowledgement; 11%, domestic acknowledgement; and 2% are below the standards required in the United Kingdom. “This represents a remarkable achievement and confirms that we’re one of the main global research powers,” declared David Eastwood, chief executive of the Higher Education Funding Council for England (Hefce), one of the agencies in charge of the evaluation, when he announced the results. “Out of the 159 institutions, 150 are doing some cutting-edge work in global terms.”

Though the RAE 2008 does not release a consolidated ranking of the institutions, an analysis of the data conducted by Times Higher Education shows that the best performances were from the universities of Cambridge and Oxford, followed by the London School of Economics and by Imperial College. Some institutions improved their performances relative to the preceding RAE, conducted in 2001; this is the case of the University of London- Queen Mary College, which climbed from the 48th to the 13th place. Other universities dropped, however; this is the case of Warwick, which slipped from 6th to 9th place. However, the universities will have to wait until March 4 to find out exactly who will gain and who will lose funding, because the division is to take into account not only research quality, but the volume of researchers submitted to assessment from each institution. The University of Cardiff, for instance, slid from 8th to 22nd place, but most probably it will not lose funding because a larger number of its faculty took part in the RAE than in

PHOTOS: WARWICK UNIVERSITY

UNIVERSITY OF LEEDS



the preceding evaluation. Even so, shock waves are expected at those institutions where performance fell, in the form, for instance, of dismissals, as happened in previous evaluations.

The RAE is distinguished by the sophistication of its methodology, based on a peer evaluation system that involves both domestic and foreign consultants, and by its magnitude – it cost US\$ 17 million, vs. US\$ 8 million for the 2001 evaluation. It is based on 15 panels that supervise the work of 67 sub-panels. Overall, 950 reviewers took part in the process. There is at least one foreign researcher on the main committees. “The idea is not to compare the assessment of international members with that of domestic members, but to ensure that the levels of quality required are the appropriate ones,” said Ed Hughes, the RAE 2008 manager from Hefce, to Pesquisa FAPESP. “In many cases, the international members help to establish parameters. They play an important role, ensuring that the panels’ analyses have international credibility.”

**Consequences** - For the purposes of comparison, the British model has significant differences relative to the system conducted in Brazil by Capes (the National Coordinating Office for the Upgrading of Personnel with Higher Education), which has been evaluating Masters and PhD programs since the 1970s, starting with the purposes and consequences of the evaluation processes. In the Brazilian case, the triennial assessment of the masters, professional masters and PhD programs aims not only at measuring the quality of the programs but also at encouraging their development, given that it provides guidance for the financing of grants and acknowledgement of the excellence of the respective research groups. Closing courses that receive a poor evaluation only happens in extreme circumstances and programs with regular grades maintain the right to train masters and doctors, even though their prestige may be affected. The RAE, on the other hand, has an immediate and sometimes devastating effect, which may extend beyond research and post-graduate studies, because

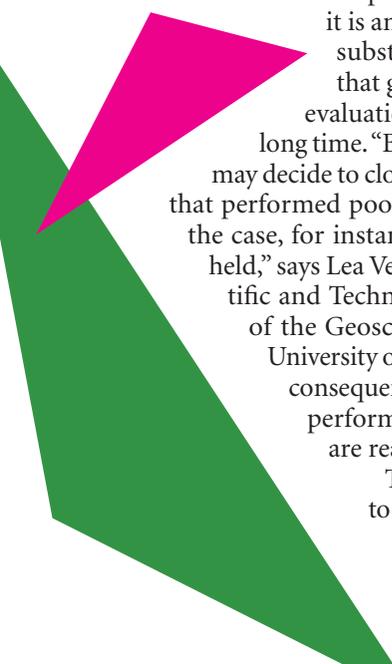
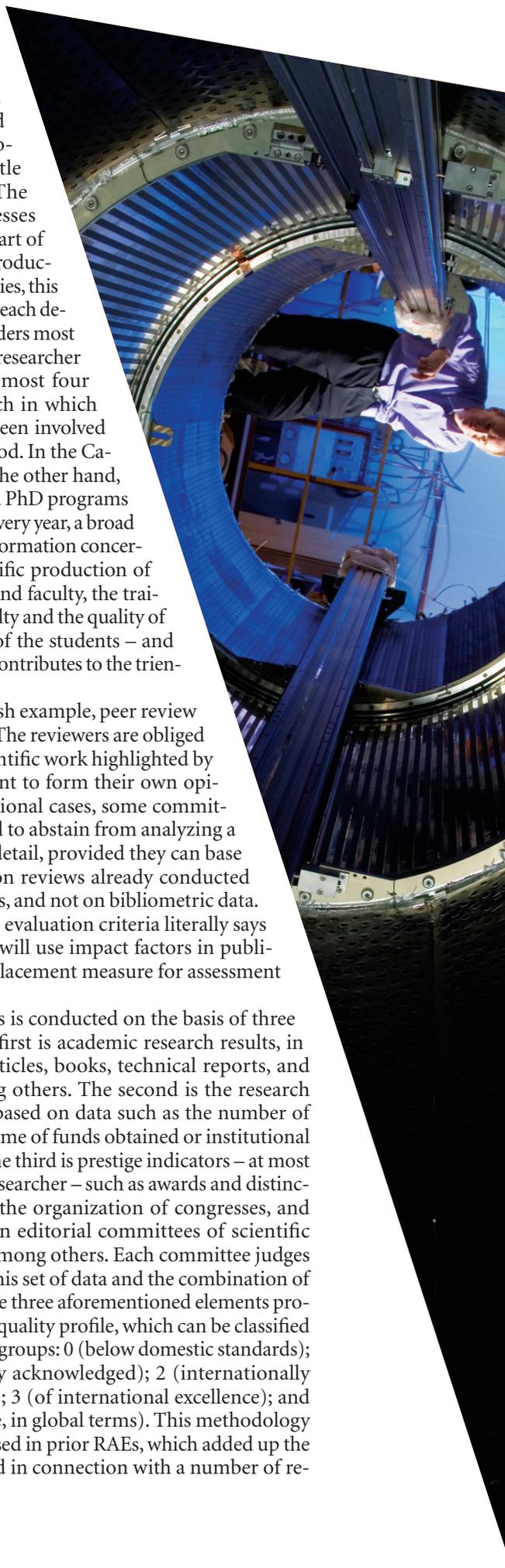
it is an input for the allocation of a substantial part of the resources that go to UK universities. A poor evaluation results in less money for a long time. “Based on the RAE, universities may decide to close down certain departments that performed poorly in the evaluation, as was the case, for instance, when the first RAE was held,” says Lea Velho, a professor at the Scientific and Technological Policy Department of the Geosciences Institute of the State University of Campinas (Unicamp). “The consequences for the departments that performed poorly in the evaluation are real,” she states.

Though both models take into account quantitative data and

peer evaluation, the Capes and Hefce methodologies have little in common. The RAE only assesses the quality of part of the academic production of universities, this part being what each department considers most relevant. Each researcher can declare at most four lines of research in which he or she has been involved during the period. In the Capes model, on the other hand, the masters and PhD programs must provide, every year, a broad spectrum of information concerning the scientific production of both students and faculty, the training of the faculty and the quality of the education of the students – and this set of data contributes to the triennial evaluation.

In the British example, peer review is the keynote. The reviewers are obliged to read the scientific work highlighted by each department to form their own opinion. In exceptional cases, some committees are allowed to abstain from analyzing a given work in detail, provided they can base their analysis on reviews already conducted by other experts, and not on bibliometric data. The chapter on evaluation criteria literally says that “no panel will use impact factors in publications as a replacement measure for assessment of quality.”

The analysis is conducted on the basis of three elements. The first is academic research results, in the form of articles, books, technical reports, and patents, among others. The second is the research environment, based on data such as the number of grants, the volume of funds obtained or institutional research aid. The third is prestige indicators – at most four for each researcher – such as awards and distinctions granted, the organization of congresses, and participation in editorial committees of scientific publications, among others. Each committee judges the quality of this set of data and the combination of the results of the three aforementioned elements provides a general quality profile, which can be classified into one of five groups: 0 (below domestic standards); 1 (domestically acknowledged); 2 (internationally acknowledged); 3 (of international excellence); and 4 (cutting-edge, in global terms). This methodology replaced that used in prior RAEs, which added up the points obtained in connection with a number of re-



quirements. "The aim is to avoid repeating distortions in the distribution of funds, with a department being ranked as 5\* getting far more money than another one also ranked as 5, although the difference between them may be small," said Ed Hughes.

In the case of Capes, the bibliometric criteria carry a lot of weight, even though evaluation is the responsibility of committees of experts. The main scientific journals were ranked by the agency according to their quality (meaning impact factor) and circulation reach (local, domestic and international). This system, called Qualis, is used to evaluate researchers' scientific articles and provides the foundation for a substantial portion of the evaluation process, especially in those areas whose academic output is expressed in articles published in journals. Thus, a modest production published in high impact publications carries more weight in the formulas used by the evaluation committees than a larger production published in periodicals with a more limited impact. The data collected are submitted to the area committees and each one of them uses specific criteria for analyzing the information. The programs are given a grade from 1 to 5. This work produces spreadsheets, common to all programs, that aim at providing transparency and that force the committees to take into account a standardized series of information such as the faculty contingent, the number of theses and dissertations submitted, articles published in national and international scientific periodicals, work published in the proceedings of national and international events, books and book chapters. However, a qualitative analysis may be required regarding topics such as the evaluation of books or book chapters, more common in the output of the human sciences, given the lack of indicators for evaluating their quality.

The doctoral programs that achieved the maximum grade (5) may be submitted to a second evaluation stage, of a more qualitative nature. They can then be re-evaluated and graded 6 or 7, depending on indicators such as the capacity to generate research groups of international standing, as measured by criteria such as the existence of international agreements, the presence of visiting professors from foreign universities generally regarded as first rate, the interchange of students with foreign universities, and the participation of faculty in committees and in the executive offices of international associations, among others.

The selection of the evaluators is also different in the two models. In the case

of the RAE 2008, there was a certain competition to fill the positions of members and heads of panels and sub-panels. These people were chosen by the representatives of the funding agencies based on 4,948 nominations provided by 1,371 scientific societies and institutions (the universities were not allowed to propose members). At Capes, the committee coordinators, who are chosen by the institution, have a degree of freedom to suggest with whom they plan to work, subject to compliance with the criteria of competence in the field. In any event, the names must be approved by the agency's executive office for evaluation. The latest triennial evaluation involved some 700 reviewers. At least 50% of the members of each committee must be replaced every three years.

**Citations** - The RAE 2008 will be the last British evaluation to follow this model. To cut costs and make the assessment faster, the UK government has decided to introduce a new system, the Research Excellence Framework (REF), which, even though it will not abandon peer evaluation, will rely heavily on bibliometric indicators, such as the number of citations of scientists' publications. "The elements that will be used and the balance between them will vary in accordance with the characteristics in each field of knowledge," states Ed Hughes. Hefce is conducting a pilot study involving 22 fields of knowledge to compare the results of the RAE 2008 with the future REF methodology. The change has split the opinion of the British scientific community, mainly because it is still unclear what methods will be used. "Taken in isolation, citations have repeatedly proven to be a poor measure of research quality," stated an editorial in the journal *Nature*, January 1st edition, regarding the changes.

The journal mentions a 1998 study that compared the results of two analyses of a set of articles on physics; one relied on metrics such as citations, whereas the other was based on peer review. The divergences reached 25% of the articles analyzed. "The policy formulators have no other option but to acknowledge that peer review plays an indispensable role in the evaluation," stated *Nature*.

In a report presented in 2003 to Hefce, researchers Nick von Tunzelman, from the University of Sussex, and Erika Kraemer-Mbula, from the University of Brighton, informed that, despite criticism of the system of evaluation by British peers, only very few countries resorted to purely quantitative systems to evaluate research and, wherever this had occurred, as in Flanders (Belgium), the measure was considered highly controversial. The issue, according to Ed Hughes, is finding the right balance. "The new system will keep some elements of peer evaluation, but we must find a way to produce a simpler and more efficient evaluation system without losing the value derived from the RAE's strict methods," he stated. ■