

The Editorial that introduces this second issue of Substantia focusses on a critical issue, the assessment of the quality of scientific research.

An everlasting question that is currently debated in several institutions is: how can we measure and evaluate the performance of individual researchers? This is a terrifically concrete and overwhelming process that affects and afflicts most of scientists in their career, and ultimately the progress of Science. The outputs of Science are not manufacturing industrial products. We need to speak out about the use of more or less obscure algorithms whose final result becomes a judgment on the quality of research. The use of algorithms does not guarantee impartial neutrality, as claimed. It is typical of those who cannot really and deeply evaluate, and therefore must rely on numbers: impact factors, h indexes, citations, etc. This Editorial tries to contribute to the current international debate by presenting the case of the Italian National Agency ANVUR.

I also recommend you to read a nicely acute and amusing paper written by Gregory A. Petsko for Genome Biology in 2008, entitled "Having an impact (factor)". Not only for fun, but for your thoughtful consideration.

24/09/2017

Pierandrea Lo Nostro

Editorial

How do we recognize a good scientist?

On June 11, 2010 before actually starting its activities, ANVUR - Agenzia Nazionale per la Valutazione del Sistema Universitario e della Ricerca (*National Agency for the Evaluation of the University and Research System*) had raised positive expectations in Italy. It was quite clear to everyone, inside and outside the academic community, that it was useful to carefully check the quality of both teaching and research activities, respectively in all Italian universities and Research Institutes.

In these seven years, ANVUR has released several reports, and has often expressed its assessments of the quality of both University education and scientific research. However, there is no longer a general support of evaluation. ANVUR's practical activity has aroused consensus, as well as criticism, both inside and outside the academic world.

Several people criticized the way in which assessments have been used. Indeed, the Italian governments of the last seven years, have allocated remarkable "reward shares" out of the scarce FFO (Fondo di finanziamento ordinario - *Ongoing Financing Fund*) to those universities which had been the best, according to ANVUR's judgement. This has meant a flow of funds from the Universities of Southern Italy up to those of the Centre and North of the country. As a consequence, the University system of the most economically depressed area of Italy has been further weakened, whereas the richer northern areas have taken advantage of it. In turn, fewer young people have enrolled in the South. Also, many boys and girls from the South moved towards the

universities of Central and Northern Italy. Many other youth did not enroll at all. However, ANVUR cannot be blamed for this usage of the evaluation of university system. Nor the idea of evaluation itself can be blamed. In fact, this is a political responsibility.

This choice does not concern Italy only: it is actually a general problem. The idea is spreading all around Europe and the whole world, that research is an enterprise like any other, and that human and financial resources should be concentrated in a few centres and universities of excellence, which can "compete and win" on the International market for the production of knowledge and education. From this perspective, most research institutes and universities only fulfil residual tasks.

This choice is a direct attack against both democracy and knowledge, as well as against the effectiveness of scientific research. It is indeed an attack against the democracy of knowledge, since it involves the fact that only a chosen few can get access - as researchers, lecturers or students - to universities and centres of excellence. In the case of students, there is a clear inequality. In several countries - from the United States to the United Kingdom - the tuition fees for universities of excellence are so expensive (tens of thousands euros a year for foreign students in some British universities) that only the children of very rich families can enrol. On the other hand, education is not a "rival" good, which diminishes as it is used. Quite the opposite: the more it is used, the more it grows. This was clear to

Vannevar Bush, as he wrote his report to the US President in 1945 – *Science, the Endless Frontier*, the “manifesto” of modern science policy. In it, Bush stated the need to enlarge the recruitment of “brains” necessary for the scientific development, by opening the university doors to the children of all US families, because intelligence does not belong to one social class only, but is rather transversal by definition. During the Second World War, the U.S. were planning to take – thanks to scientific research too – the economic and cultural leadership of the planet. In the same way, the whole world needs the intelligence of everyone.

Another kind of criticism has been made by a section of the Italian academic community, and concerns ANVUR’s method of evaluation. Indeed, they say ANVUR uses in an excessively overriding and rigorous way the typical parameters of bibliometry when assessing the research quality of a university: namely, the number of published papers; the impact factor of journals in which they are published; and, finally, the number of quotations of each paper.

This type of criticism has a wider significance. Indeed, many people all over the world are wondering whether the bibliometric method may indeed be considered the best evaluation method for research, or even for researchers.

It is true that, in a world where several millions people devote their lives to science, in the framework of a growing number of International projects, it is convenient to find a universal evaluation method for research activity. However, it is also true that, if we reduce evaluation to the mere analysis of bibliometric parameters, this may produce misleading results.

Indeed, bibliometric analysis has its own inherent limits, as pointed out by an endless scientific literature. We cannot analyze them in detail here. We’ll just consider its main limits, since they can steer the evolution of the International research and Higher Education system in undesirable directions.

In bibliometric analysis, research quality and quantity tend to match. Normally, quality is assessed by measuring quantity. Now, the number of papers published in International scientific journals, as well as quotations obtained, are significant indicators of a researcher’s talent. However, they are not the only indicators, and probably not the main ones.

In any case, the evaluation of quality as based exclusively on bibliometry is not only incomplete: it is also misleading.

Young people, in particular, pay a high price for this evaluation system. First of all, even geniuses are penalised, because they have had no time as yet to publish a lot of papers and get quotations. Secondly, this leads to a

vision of scientific activity as based upon the “publish or perish” principle, rather than upon good ideas.

Bibliometric evaluation, when used in the wrong way, may become a levelling-out power in the research community, for more than one motive.

First of all, the “publish or perish” principle tends to eat away at scientific creativity, in favour of Thomas Kuhn’s “normal science”. Even the search for a high number of quotations may become a levelling-out element, because it leads to join “fashionable” schools of thought, rather than look for originality, which is one of the five values considered by Robert Merton as the bases of scientific enterprise.

The exclusively quantitative pressure becomes a levelling-out power, not only for individual researchers, and for small groups of researchers, but also for large Institutes and broad areas of science. Indeed, both financial and human resources tend to be concentrated in those Institutes and areas which are better assessed. As a consequence, small – but promising – institutes and areas suffer from a worse assessment – as in the case of the flow of students from Southern to Northern Universities in Italy. As a result, we may get a world scientific system constituted by a sea of mediocrity and a few islands of excellence, where many papers are published and lots of quotes are made, where as Kuhn’s “dominating paradigms” are not challenged.

It would be something of a paradox that an age so rich in scientists – the world has never had so many of them – also becomes an age poor in groundbreaking scientific ideas.

Hence the need, which is also felt outside academia, to get past the bibliometric method and look for a satisfying answer to the difficult question, which the German physicist Reinhard Werner recently posed in *Nature*: «How do we recognize a good scientist?»

However, the pooling of human and financial resources in few “knowledge firms” competing on the International market, together with the wrong method of evaluation of research quality can lead to the end of science itself. This has already happened in the past: Hellenistic science, for instance, was “forgotten” with the Romans conquering the Mediterranean area. It took Europe a millennium and a half in order to go back to its own true nature. In the same way, if science resorts to seeking – like any other firm – immediate results which may increase its competitiveness; if it entrenches itself smugly in a few fortresses; if it promotes uniformity rather than innovation, it risks dying out.

Therefore, the urgent question now is: “How do we recognize and save a good science?”