An Analysis of Scientific Research Performance in Italy: Evaluation Criteria and Public Funding

Gennaro Guida[1]

[1] Department of Economic and Legal Studies, University Parthenope, Naples, Italy

Correspondence: Gennaro Guida, Department of Economic and Legal Studies, University Parthenope, Naples, Italy. E-mail: gennaro.guida@uniparthenope.it

Abstract

In the last ten years, the assessment of scientific research has been useful for two main reasons: first to provide researchers with a shared methodology able to assess scientific productions and second to enable governments to locate investments where they yield the best results. The aim of this article is to investigate the methodology used to assess academic performance in Italy, focusing on the one hand on its application to economic sciences and on the other on the way in which it influences the distribution of the budget and the future performance of an institution. In this regard, an analysis of sample data extracted from the well-known Scopus database raises doubts about the advisability of linking the distribution of funds to the evaluation of the quality of the research.

Keywords: bibliometrics, citations, public administration, research policy, scientific research

1. Introduction

In Italy, over the last ten years, great importance has been attached to the new agency created to assess scientific research. This agency, the ANVUR (National Agency for Assessing University and Scientific Research) was established in 2006 to perform the following functions:

1) External assessment of the quality of activities of universities and public and private research organizations receiving public funding on the basis of an annual programme approved by the Ministry of Education, University, and Scientific Research;

2) Addressing, coordinating, and monitoring the evaluation activities carried out in the internal assessment cores of universities and research entities;

3) Evaluation of the efficiency and effectiveness of state funding programmes and of the incentives to undertake research and innovation activities.

As written in Italian law, "The results of ANVUR's evaluation activities constitute a reference criterion for allocating state funding to universities and research entities".

Due to the importance of the intended purpose and the delicate nature of the assignment, the legislator wanted to emphasize the principles on which the agency is founded: independence, impartiality, professionalism, and transparency. Through the Assessment of Scientific Research (VQR), the ANVUR measures the validity of the results of the scientific research carried out in a period of the past three years by state and non-state universities and private entities that engage in research activity. On the basis of the results of the VQR, state funding is allocated to universities and research entities. The following work will present a summary of the issues related to the evaluation of research, the legislative choices made in Italy, the methodology chosen for the distribution of funds and the problems that this may generate, in contrast with the university's educational needs. The article is divided into sessions: the following session will discuss about the known problem on the choice of evaluation criteria to be adopted, the third will present the choice of criteria carried out by ANVUR for the evaluation of research in the economic area, the fourth will expose the mechanism for allocating state funding to universities, the fifth session will analyze a sample of data replicating the evaluation with bibliometric indexes to identify the most productive areas, comparing them with those where university's educational needs are higher, and finally the conclusions.

2. Peer Review or Bibliometric Indexes? A Review of the Literature

The assessment of scientific research has certainly not arisen today but has deep historical roots, even in times when it was implemented without specific formalization. It suffices to think of the evaluation undertaken by experienced university professors of the work conducted by their graduates/assistants, determining a disadvantage or advantage for their careers in scientific research.

This type of approach, although widely debated and certainly more formalized, is still valid today and is called peer review. It is the most widely used national and international qualitative tool for evaluating research. According to Alberto Baccini (2010), peer review can be defined as "a set of heterogeneous and non-standardized practices through which a group of individuals expresses a judgment on the other's scientific works to determine their quality. Professors called to express such a judgment are selected by a large group of auditors who are considered equal to those who produced the job to judge". This methodology has some positive aspects: the guarantee of a specialized team capable of assessing a scientific production, for example, gives the readers the perception that they are reading something excellent, because it was able to pass a specialized assessment exam. At the same time, however, peer reviews can often be questioned and not shared by those who are subjected to them, because they are based on subjective assessment and in rare cases they can sometimes generate fraud and suspicion (see discussion in Sandstrom & Hallsten, 2008). Another criticism of peer reviews is that auditors tend to have a more magnanimous attitude towards authors with particular characteristics (Peter & Ceci, 1982; Garfunkel et al., 1994; Hodgson & Rothman, 1999); for instance, belonging to a prestigious American or English university can favour the acceptance of articles submitted to scientific journals regardless of the scientific content. Moreover, peer review times are usually quite long, the process taking several weeks or months. The most critical opponents consider it as expensive, slow, prone to bias, open to abuse, possibly anti-innovatory, and unable to detect fraud (Pouris, 1988; Smith, 1997; Wenneras & Wold, 1997) .

All these problems have laid the foundations for the formation and diffusion of a new science, bibliometry.

Bibliometry is a science applied to various scientific areas and uses mathematical and statistical techniques to analyse the distribution patterns of scientific publications and verify their impact within scientific communities. It is the first measuring instrument that we can define as quantitative, and it accompanies peer reviews (a qualitative instrument) and in some cases replaces them.

The first traditional instrument used in bibliometry is the analysis of citations to publications that a scientific article receives from other articles. The higher the number of quotations an article receives, the more importance its contents will probably have.

The use of bibliometry in the assessment of research has taken place since the second half of the last century. Eugene Garfield, a US chemist and businessman who founded Eugene Garfield Associated in 1954, established the Institute for Scientific Information (ISI) in 1964 with the aim of creating a useful tool to offer assessment services to institutional bodies, as much in the public sector as in the private sector. In 1992 Garfield sold the ISI to the famous Thomson Reuters multinational. Since 1964 the ISI has become one of the most important and prominent bibliometric service providers: a real agency offering consultancy with offices all around the world. The work carried out by the ISI has been made possible by the creation of particular databases that collect data and metadata from selected magazines. These databases vary in relation to the disciplinary field to be analysed.

The remarkable success of bibliometry in recent years is linked to the identification of objective parameters for measurement, called bibliometric indicators, aimed at assessing research and researchers. Scientific journals specializing in the topic of bibliometry and citational analysis have also emerged.

The review of the most-used bibliometric indexes in the literature inevitably involves the one generated by its founder, Garfield, known as the impact factor (IF). The IF measures the average number of citations received in a particular year from articles published in a scientific journal in the previous two years. The main strength of this index is its simplicity of use for the evaluation of scientific research. This index, it should be emphasized, is intended as an index of evaluation of the journal in which an article is published and not as an evaluation of the research of a scholar. Despite being the most widely used index, it is not exempt from criticism. For example:

1) The use of the impact factor in assessing research is related to the citational behaviours of the various scientific communities. This makes it less homogeneous if used to compare the productivity of two different research fields;

2) The type of publication has a weight: less specialized journals, or those with many methodological contents, are generally cited more than ultra-specialized and experimental articles. The same is true of magazines focusing mainly on the publication of synthesis articles (reviews, etc.); even though no original contribution

is made, such magazines tend to be cited frequently, because researchers use the reviews, especially those written by prestigious colleagues, such as syntheses of the previous literature;

3) Seglen (1997) highlighted that the impact factor is the result of an average score associated with a magazine. In fact, the same magazine might contain articles judged positively and articles criticized negatively by critics: in other words, articles with strong citations and articles with an inconsistent or no impact. The formula of the impact factor does not show these differences among the publications of the same magazine.

Arguments that at different times have challenged such an index have produced a development in scientific assessment so that today more shared metrics are available for the assessment of scientific research.

Hirsch's index was created in 2005, adopting the name of the Californian scientist who constructed it at the University of San Diego (also known as the h-index or index h) by publishing "An Index To Quantify an Individual's Scientific Research Output" in the open access *ArXives* magazine.

The h-index is defined as follows: a scientist has an index h if at least h of his Np jobs have at least h quotes each and the remaining Np-h jobs have at the most h-1. The h-index compares authors belonging to the same discipline, and unlike the IF it does not make comparisons between authors from different scientific fields, which often have no point in common. It also takes into account the age of the researcher and the number of years during which he has published his work. Indeed, the h-index is able to measure the impact of the researcher within his discipline using the number of publications issued and quotations received.

Table 1. H-index

PRO	CON
The index assesses the authors belonging to the same discipline to make comparisons.	The index does not make comparisons between authors belonging to different scientific fields.
It takes into account the age of the researcher and the number of years during which he has published his work. It does not concern the importance of the scientific journal in which an article is published, as it aims only to measure the impact of the scientist in his academic field.	It does not consider the entire volume of publications of a scientist but only those that have had greater success.
	It does not exclude quotes from the same author.
	It tends to disrupt younger researchers or those who have just started researching: the h-index is more effective from a long-term perspective.
	It allows scientists who have already created a position to maintain a high index, as it does not decline over time, even if no work has been published for years.

The following table provides a list of bibliometric indexes that are less famous than the impact factor or the h-index but were built with the aim of improving them or proposing shared alternatives. Some of them are used for the assessment of Italian universities (AIF, SJR, IPP, and SNIP).

Table 2. Other common bibliometric indexes

INDEX'S DENOMINATION	HOW TO DERIVE
Immediacy Index	The index measures how successful an article is in the year of publication and how fast a magazine is cited. It is calculated by dividing the current citation number of a magazine in a given year by the number of articles published in the current year.
M-Index	It is obtained by dividing the h-index by the number of years of academic activity (considering the beginning of the scientific activity as the release of the first publication).
Scimago Journal Ranking	This is a measure of scientific influence of scholarly journals that accounts for both the number of citations received by a journal and the importance or prestige of the journals that contain such citations. The SJR indicator computation is carried out using an iterative algorithm that distributes prestige values among the journals until a steady-state solution is reached. The SJR algorithm begins by allocating an identical amount of prestige to each journal then using an iterative procedure; this prestige is redistributed in a process whereby journals transfer their achieved prestige to each other through citations. The process ends when the difference between journal prestige values in consecutive iterations no longer reaches a minimum threshold value.
Source Normalized Impact per Paper	It measures the impact of a citation based on the total number of citations in a given disciplinary area: the impact of a single citation is given a higher value in areas in which citations are scarcer and vice versa.

Impact Per Publication	The IPP corresponds to the average number of citations received in a year from articles published in the magazine over the past three years.
Eigenfactor	The Eigenfactor score is a rating of the total importance of a scientific journal. Journals are rated according to the number of incoming citations, with citations from highly ranked journals weighted to make a larger contribution to the Eigenfactor than those from poorly ranked journals. As a measure of importance, the Eigenfactor score is scaled with the total impact of a journal. All else being equal, journals generating a greater impact in the field have larger Eigenfactor scores.
Article Influence Score	The Article Influence Score measures the average influence of articles in the journal and is therefore comparable to the traditional impact factor.
G-index	The index is calculated based on the distribution of citations received by a given researcher's publications, such that, given a set of articles ranked in decreasing order of the number of citations that they received, the g-index is the unique largest number, such that the top g articles received together at least g^2 citations.

The pros and cons of peer-review and bilbiometrics methods have been thoroughly dissected in the literature (Horrobin, 1990; Moxam & Anderson, 1992; MacRoberts & MacRoberts, 1996; Moed, 2002; van Raan, 2005; Pendlebury, 2009; Abramo & D'Angelo, 2011).

On the dispute about which is the best choice to evaluate scientific research we mention the works of two professors, whose ideas feel very close and that we share. Butler (2007) defines the boundaries of the use of bibliometry by noting that 'metrics' have their place, and can make the process more efficient and cost-effective, but peer review must be retained as a central element in any research assessment exercise. The role of metrics is as 'a trigger to the recognition of anomalies', rather than as a straight replacement for peer review. A real challenge is to combine the two methodologies in such a way that the strength of the first compensates for the limitations of the second, and vice versa (Moed, 2007). The combined use of the two methods is the most shared tool.

3. Methodology to Assess Italian Scientific Research in Economics

In Italy scientific research in economics is assessed by a group of experts (called the GEV) chosen by the Ministry. The GEV's evaluation of the products follows the method of the informed peer review, which consists of using different evaluation methods, possibly independent of each other, and harmonizing them within the GEV, which nevertheless has the ultimate responsibility for the evaluation. The evaluation methods used are the following: the peer review, entrusted to external auditors (normally two), who are normally chosen by two different GEV members; the GEV's direct evaluation, which involves a peer review carried out within the GEV in the same manner as the peer review commissioned by external auditors; finally the bibliometric analysis, conducted using the indicators and algorithms defined below.

The reviewers chosen by the Ministry will be selected from among the most authoritative and scientifically qualified scholars and specialists in the disciplines examined. The external or internal auditors' assessment of the GEV is based on a special auditor's report and guidelines prepared by the GEV. The review will allow the auditor to assign a score for each of the three evaluation criteria established, namely originality, methodological rigour and attested or potential impact, and will include a free field allowing for a limited number of words; it is compulsory to provide a brief summary of the reasons for the evaluation. The GEV transforms the information contained in the auditor's report into one of the available classes of merit.

Altogether 50% of the material subjected to evaluation will be judged by bibliometric methods. Following this methodology, as is known, the analysis of the journals determines the assessment of the papers published in the journals. The GEV will collect impact indicators for 2014 from the databases ISI WoS (in particular, IF, IF5Y, and AIS) and Scopus (in particular, IPP, SNIP, and SJR). For all the journals belonging to the list, the GEV will also collect the h-index from Google Scholar for the 2010–2014 period. Magazines with a missing or zero value for the h-index will not be included in the final list.

Finally, the classification procedure will ensure that the ex ante chance of each article falling into one of the assessment classes is defined as follows:

Excellent [the top 10% of the distribution of international scientific production in the area to which it belongs];

High [10%−30% of the distribution of international scientific production in the area to which it belongs];

Discreet [30%−50% of the distribution of international scientific production of the area to which it belongs];

Acceptable [50%−80% of the distribution of international scientific production in the area to which it belongs];

Limited [80%−100% of the distribution of international scientific production of the area to which it belongs].

4. Resource Allocation Following the Assessment of Scientific Research

Italian law sets out the criteria for allocating 20% of the resources destined for the purposes of rewarding the universities with the following percentages: 65% based on the results of the Assessment of Scientific Research (VQR 2011−2014); 20% based on the Evaluation of Recruitment Policies (VQR 2011−2014); 7% based on the results of teaching with specific reference to the international component; and 8% based on the learning outcomes with specific reference to the number of regular students who have passed exams.

This means that considerable attention is paid to the outcome of scientific assessment (as stated, 65% of the weight in the allocation of state funds). The output comes from two operations above all: the first operation calculates the index "A", which is capable of assessing products of each scientific area of research. This index is a weighted average of five indexes able to measure the following aspects:

1)   With a weight of 0.75, the ratio between the sum of the evaluations obtained from the products presented by the institution in the area and the overall area evaluation;

2)   With a weight of 0.20, the ratio calculated as previously in the subset of research publications and products submitted by research workers who, during the 2011−2014 assessment period, were recruited by the institution or incardinated in a higher band or role.

3)   With a weight of 0.01, considering the funding obtained from participating in competitive bids for national or international research projects.

4)   With a weight of 0.01, considering the number of doctoral students or students enrolled in medical and health care specialization schools, research grant holders, and post-docs.

5)   With a weight of 0.03, the performance variation compared with the VQR of the previous period.

The second index is based on the performance of the university, and it is calculated as a weighted average of sector indicator A for each area. This is the final index of the performance of an institution.

Ben Martin and Aldo Genua (2003) analysed, in a not recent but accurate work, the methodologies present in the developed countries that are capable of comparing the most widely used techniques for assessing scientific research and identifying the different purposes that each country attributes to it. A shared use of peer reviews emerges as the main and most widespread research methodology. Accordingly, the scientific community implements self-regulation mechanisms and proposes defences against attempts to introduce external-logical evaluations that often push research into unknown domains. The differences found regard two aspects:

1). Focusing solely on the research product or considering the entire production process of scientific research;

2). The links between assessment and public funding to reward scientific research.

In this regard, we can compare two European experiences, the English RAE (Research Assessment Exercise, the old English methodology, which inspired the Italian VQR) and the Dutch SEP (Standard Evaluation Protocol). In the first case, from 1986 until the RAE was replaced by the current evaluation system (REF Research Excellence Framework, 2013), the evaluation exercise was primarily intended to provide a comprehensive assessment of the work of the national academic sector through an ex post evaluation of the research products (four for each researcher), using the peer review method. On the basis of the results of the UK assessment exercise, funds were distributed to universities for the next two years at the end of the fiscal year (e.g. RAE 2006/2008 allocated funds automatically for the two-year period 2009-2010).

The Standard Evaluation Protocol 2009-2015 provides for an internal self-assessment followed by external evaluation based on peer reviews. The assessment concerns both the research facilities and the research programmes on which the work is carried out. The evaluation is performed on a disciplinary basis and concerns the quality of research products, management, research policies, and doctoral training. The result is a rating assigned to each structure in relation to the quality of the publications, the productivity of the research structure, the social relevance of the research results, and the prospects for developing the research domains on which it operates. The outcome of the assessment has no consequences for the transfer of funds to universities.

These examples are useful to understand the future direction for the developed countries. On one hand, there is a system that entrusts the future of the academic world, tout court, to the assessment of scientific research as the output of this world, following the idea of value for money. In this case, the financial instrument will decide the future capacities of an academic institution. The academic world will probably converge towards more prestigious and large-scale structures. On the other hand, there is a mechanism capable of identifying the

weaknesses in the university system, without the incentive of a pricing mechanism, it could find difficulties for solving problems. The future debate will consider these problems again.

5. A Survey of Italian Universities

The Italian method of assessing scientific research uses state funding as an incentive to improve productivity. The risk underlying the policy line adopted is twofold. On the one hand, the academic world and the scientific research are subjected to an evaluation that, regardless of the criteria adopted, makes it dependent on them, and the researcher's aim could shift from conducting good research to obtaining a good evaluation. On the other hand, one wonders what would happen to the size of the universities and their presence in the territory if the competitive advantages of the most productive ones were to increase.

What is the situation of the productivity of Italian scientific research in qualitative and quantitative terms? To answer this question we will use a sample extracted from the available databases. We should first choice the source of data. Scopus and Web of Science are two world largest abstracting and citation databases. Publishing Corporation Elsevier currently owns Scopus. Scopus covers 69.2 million documents. Clarivate Analytics currently owns Web of Science. Its key database, Web of Science Core Collection, contained 66.7 million documents. Scopus and Web of Science have their advantages and limitations for the different aspects of the quantitative scientometric analysis (e.g. Deis & Goodman, 2005; Bakkalbasi et al., 2006; Meho & Yang, 2006; Falagas et al., 2008; Archambault et al., 2009; Vieira & Gomes, 2009; Li et al., 2010; Shashnov & Kotsemir, 2015; Kotsemir & Shashnov, 2017). A key advantage of Scopus for our analysis is the presence of the system of unique author and organization identifiers (profiles). The user, then, can analyze the publication activity of organizations, while in Web of Science this chance is not easily available.

We can analyse a sample of about 4,000 articles published by professors of economics employed in Italian universities, extracted from Scopus database, to which we have added information regarding some characteristics of each university, extracted from the Ministry website, as the number of students, and the values of the bibliometric indicators used for the evaluation of scientific research (the above-mentioned indexes). The choice of selecting the economic area has a twofold motivation: the territorial diffusion of the economic area sealed by the presence of the faculty of economics in many universities, and the choice of the legislator to evaluate part of the work through bibliometric indexes, a choice common to many others sectors.

Through data available, we can replicate the evaluation of the articles included in sample data and draw some interesting considerations.

The following colour maps show the educational commitments of the Italian universities of economics (number of student enrolled) for province, and the evaluation of the quality of the research, as carried out by the ANVUR.
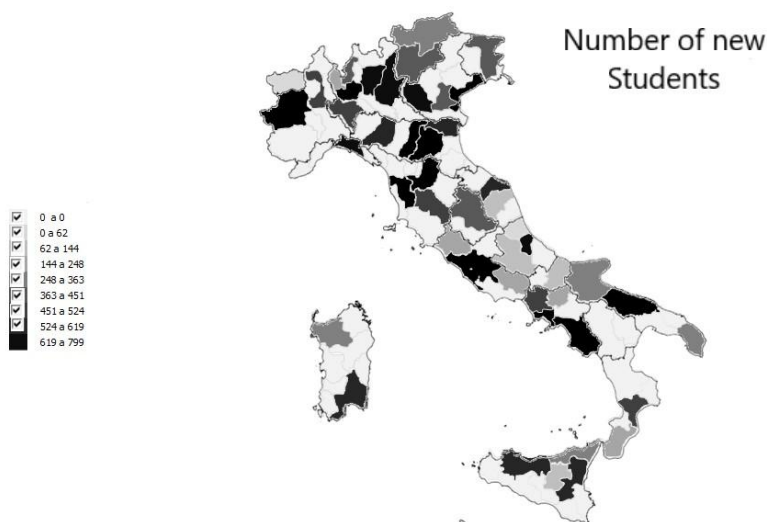

Figure 1. Number of new students

We can see from map 1 the distribution of new students for provinces. This distribution follows, as expected, the distribution of the population in Italy, highlighting a heavy load for universities in the most populous areas.

The choice of Italian legislator, inspired by the English RAE, is based on the following principle: on the basis of the results of the assessment exercise, funds were distributed to universities. By linking state funding allocation to the evaluation of scientific research we must consider the results of the VQR. From the sampled data processed, we replicate the results (for economics) with bibliometric indexes and present them in the following map, highlighting the best results by province.
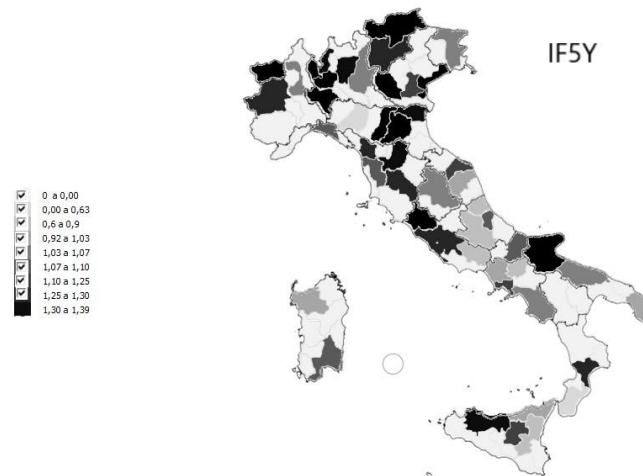

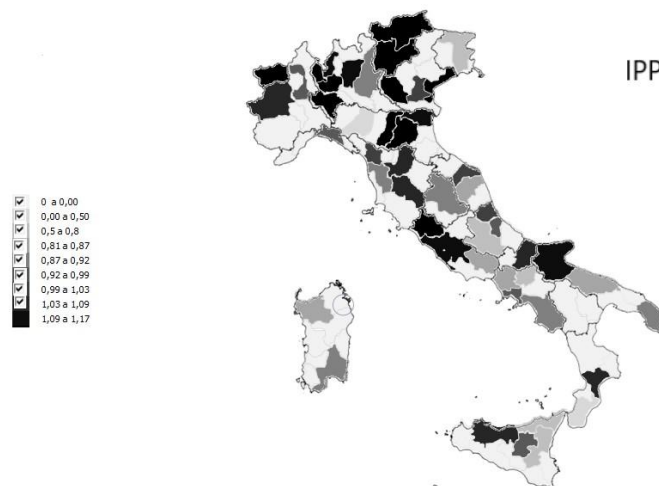Figure 2. Impact factor on five years
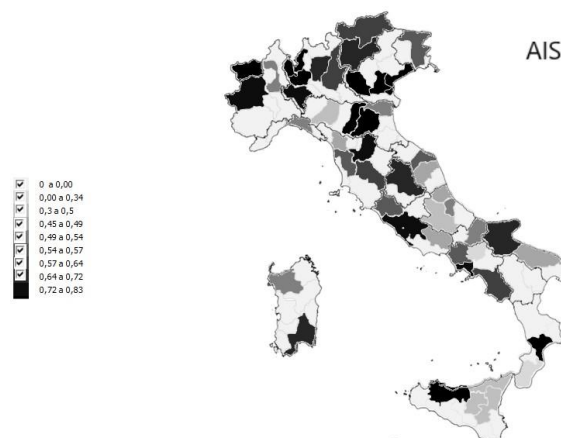

Figure 3. Impact per publication
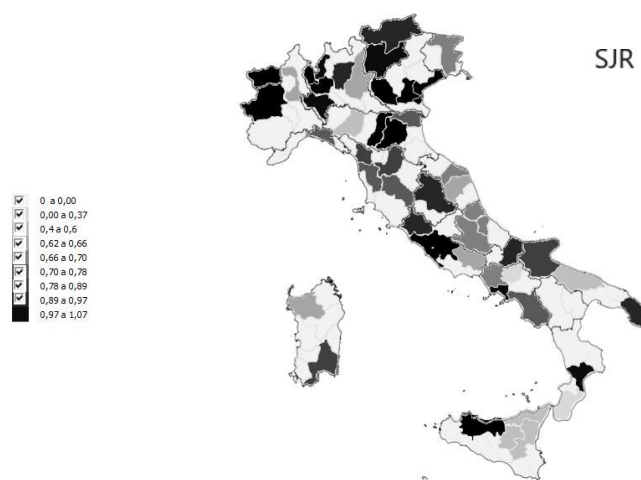

Figure 4. Article influence score

Figure 5. Scimago journal ranking

As we can see from the maps, the two main missions of the universities, teaching and research do not always coincide. If one of the Italian institutions attains a good score in the VQR exercises, and if the governance of that institution is capable of investing in human capital, then part of the funding will be spent in that area and the university will probably earn greater prestige. Researchers with high performance are likely to move there for work. However, what would happen to education and to the possibilities of citizens receiving homogeneous services throughout the country? The risk may be that research and teaching increasingly go in different directions and these differences could become more acute in a country where the percentage of adults with a tertiary education qualification (the highest level of education achieved in Italy) is among the lowest in the OECD countries, with only 18% of graduates, half of the OECD average (OECD, 2017).

6. Conclusions and Final Considerations

Making some considerations concerning the costs and benefits of scientific research assessment is a natural consequence of it, given that the resources used are substantial and that the criticism of the methodology adopted stems from multiple locations, as well as considering the importance that some countries entrust to rating.

According to a study carried out by Robert Bowman, Director of the Center for Nanostructured Media at Queen's University in Belfast, the costs of the assessment of scientific research in the UK consist of the following: 600 million pounds for the preparation of documentation to be evaluated, 300 million for their selection and validation, 200 million for centralized documentation management, and 100 million for case studies. Professor Bowman admitted that this estimate, which takes into account the total cost of professor salaries, is the top limit and challenged his colleagues to propose realistic changes to reduce the figure to below 500 million pounds. For Italy, the online magazine "Roars" estimated the cost of the VQR to be in the region of 300 million euros, while Aldo Genua and Matteo Piolatto's estimate (2016), representing 50% of the opportunity cost, was around 182 million. There are therefore good reasons to believe that the VQR represents a substantial investment from which the community must benefit adequately.

In modern democracies, in which there is a new culture of public accountability of the public operator over citizens and of value for money related to the social and economic value of public investment, governments invest where they yield the best results. Scientific research cannot be free but must be subject to this rationale. The Italian institutional framework links the distribution of funds to the evaluation of the quality of the research. Is this the right way to improve the performance of scientific research?

To answer this question two tools are necessary: evaluation criteria capable of identifying without a doubt the best performance, and an incentive mechanism, which is often identified in a financial tool.

First: are the evaluation criteria chosen for the assessment so reliable? We answer this question quoting an episode known as the "Queen's question".

The Queen of England visited the prestigious London School of Economics, and during the ceremony she asked the following question: "Why did the majority of economists not predict the financial crisis of 2008?". We recall that the failure of Lehman Brothers in September 2008 gave rise to the biggest financial crisis since 1929, and the recession involved many countries and had serious economic and social repercussions. World-class

economists were not able to foresee the crisis or to interpret what was happening. A few months later, in December 2008, the results of the Research Assessment Exercise were published. From the ranking of the different disciplines (physics, chemistry, history, etc.), the result was that "among the disciplines considered economics and econometrics are the fields where the score is maximum". Consequently, while the "queen's question" resembled a tornado, highlighting a fundamental problem in current economic research, the result of the valuation for economic disciplines was not only good but the best of all in England. This anecdote enhances the doubts of an excessive confidence in the evaluation criteria. Academic literature, as seen is always looking for better tools, and all evaluation methods present pros and cons.

Second: the overall rationale behind funding incentives is usually that if money is given to the best performers, it will most likely produce better results and give an overall incentive for better performance. Funding shifts do not strongly affect the actual practices of research, for example, publication behaviour (Albert, 2003; Behrens & Gray, 2001; Van Looy et al., 2004), may lead to unintended negative consequences especially in terms of basic research outputs (Geuna, 1999; Ziman, 1996). Finally no straightforward connection between financial incentives and the efficiency of university research output exists (Auranen & Nieminen, 2010).

These arguments cast doubts over the advisability of linking the distribution of funds to the evaluation of the quality of the research.

The analysis presented in session 5 shows that the load for universities to meet the expectations of students to enjoy a good education system could be placed in provinces other than those financed through the framework previously exposed. In a country where the rate of graduates is very low (18%) we suggest to invest in the quality of the public service trying to reduce those imbalances that the OECD report showed, rather than rewarding dubious excellences with mechanisms of doubtful effectiveness.

References

Abramo, G., & D'Angelo, C. A. (2011). Evaluating research: From informed peer review to bibliometrics. *Scientometrics, 87*(3), 499-514. https://doi.org/10.1007/s11192-011-0352-7

Albert, M. (2003). Universities and the market economy: The differential impact on knowledge production in sociology and economics. *Higher Education, 45*(2), 147-182. https://doi.org/10.1023/A:1022428802287

Archambault, É., Campbell, D., Gingras, Y., & Larivière, V. (2009). Comparing bibliometric statistics obtained from the Web of Science and Scopus. *Journal of the Association for Information Science and Technology, 60*(7), 1320-1326. https://doi.org/10.1002/asi.21062

Auranen, O., & Nieminem, M. (2010). University research funding and publication performance An international comparison. *Research Policy, 39*(6), 822-834. https://doi.org/10.1016/j.respol.2010.03.003

Baccini, A. (2010). *Valutare la ricerca scientifica: Uso e abuso degli indicatori bibliometrici*. Bologna: Il Mulino.

Bakkalbasi, N., Bauer, K., Glover, J., & Wang, L. (2006). Three options for citation tracking: Google Scholar, Scopus and Web of Science. *Biomedical Digital Libraries,* 3-7. https://doi.org/10.1186/1742-5581-3-7

Behrens, T. R., & Gray, D. O. (2001). Unintended consequences of cooperative research: Impact of industry sponsorship on climate for academic freedom and other graduate student outcomes. *Research Policy, 30*(2), 179-199. https://doi.org/10.1016/S0048-7333(99)00112-2

Butler, L. (2007). Assessing university research: A plea for a balanced approach. *Science and Public Policy, 34*(8), 565-574. https://doi.org/10.3152/030234207X254404

Deis, L., & Goodman, D. (2005). Web of science (2004 version) and Scopus. *The Charleston Advisor,* 5-21.

Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., & Pappas, G. (2008). Comparison of PubMed, Scopus, web of science, and Google scholar: strengths and weaknesses. *The FASEB Journal, 22*(2), 338-342. https://doi.org/10.1096/fj.07-9492LSF

Garfunkel, J. M., Ulshen, M. H., Hamrick, H. J., & Lawson, E. E. (1994). Effect of institutional prestige on reviewers' recommendations and editorial decisions. *JAMA, 272,* 137-138. https://doi.org/10.1001/jama.1994.03520020063017

Genua, A., & Piolatto. M. (2016). Research Assessment in the UK and Italy: Costly and Difficult, but Probably Worth It (at Least for a While). *Research Policy, 45,* 260-71. https://doi.org/10.1016/j.respol.2015.09.004

Geuna, A. (1999). *The Economics of Knowledge Production*: *Funding and the Structure of University Research*. Edward Elgar, Cheltenham.

Hirsch, J. E. (2005). An Index To Quantify an Individual's Scientific Research Output. *Proceedings of the National Academy of Sciences of USA, 102*(46), 16569–72. https://doi.org/10.1073/pnas.0507655102

Hodgson, G. M., & Rothman, H. (1999). The Editors and Authors of Economics Journals: A Case of Institutional Oligopoly? *Economic Journal, 109*, 165-86. https://doi.org/10.1111/1468-0297.00407

Horrobin, D. F. (1990). The Philosophical Basis of Peer Review and the Suppression of Innovation. *JAMA, 263*(10), 1438-1441. https://doi.org/10.1001/jama.1990.03440100162024

Kotsemir, M., & Shashnov, S. (2017). Measuring, analysis and visualization of research capacity of university at the level of departments and staff members. *Scientometrics, 112*(3), 1659-1689. https://doi.org/10.1007/s11192-017-2450-7

Li, J., Burnham, J. F., Lemley, T., & Britton, R. M. (2010). Citation Analysis: Comparison of Web of Science, Scopus, SciFinder, and Google Scholar. *Journal of Electronic Resources in Medical Libraries, 7*(3), 196-217. https://doi.org/10.1080/15424065.2010.505518

MacRoberts, M. H., & MacRoberts, B. R. (1996). Problems of citation analysis. *Scientometrics, 36*(3), 435-444. https://doi.org/10.1007/BF02129604

Martin, B. R., & Genua, A. (2003). University Research Evaluation and Funding: An International Comparison. *Minerva, 41*, 277-304. https://doi.org/10.1023/B:MINE.0000005155.70870.bd

Meho, L. I., & Yang, K. (2006). A new era in citation and bibliometric analyses: Web of Science, Scopus, and Google Scholar. *arXiv preprint* cs/0612132.

Moed, H. F. (2002). The impact-factors debate: The ISI's uses and limits. *Nature, 415*, 731-732. https://doi.org/10.1038/415731a

Moed, H. F. (2007). The future of research evaluation rests with an intelligent combination of advanced metrics and transparent peer review. *Science and Public Policy*, 34(8), 575-583. https://doi.org/10.3152/030234207X255179

Moxam, H., & Anderson, J. (1992). Peer review. A view from the inside. *Science and Technology Policy, 5*(1), 7-15.

OECD. (2017). *Education at a Glance 2017: OECD Indicators*. Paris: OECD Publishing. http://dx.doi.org/10.1787/eag-2017-en

Pendlebury, D. A. (2009). The use and misuse of journal metrics and other citation indicators. *Scientometrics*, *57*(1), 1-11. https://doi.org/10.1007/s00005-009-0008-y

Peters, D., & Ceci, S. (1982). Peer-review practices of psychological journals: The fate of submitted articles, submitted again. *Behav Brain Sci., 5*, 187-255. https://doi.org/10.1017/S0140525X00011213

Pouris A. (1988). Peer review in scientifically small countries. *R&D Management, 18*(4), 333-340. https://doi.org/10.1111/j.1467-9310.1988.tb00608.x

Sandstrom, U., & Hallsten, M. (2008). Persistent nepotism in peer-review. *Scientometrics, 74*(2), 175-189. http://dx.doi.org/10.1007/s11192-008-0211-3

Seglen, P. O. (1997). Citations and Journal Impact Factors: Questionable Indicators of Research Quality. *Allergy*, *52*(11), 1050-6. https://doi.org/10.1111/j.1398-9995.1997.tb00175.x

Shashnov, S., & Kotsemir, M. (2015). Measuring the Research Capacity of a University: Use of Web of Science and Scopus. *In 2nd International multidisciplinary scientific conference on social sciences and arts SGEM2015. Conference proceedings vol.II . book 1: psychology and psychyatry, sociology and helathcare, education* (pp. 621-628). Sofia: STEF92 Technology Ltd., 2015.

Smith, R. (1997). Peer review: Reform or revolution? *British Medical Journal, 315*, 759-60. https://doi.org/10.1136/bmj.315.7111.759

Van Looy, B., Ranga, M., Callaert, J., Debackere, K., & Zimmerman, E. (2004). Combining entrepreneurial and scientific performance in academia: towards a compounded and reciprocal Matthew-effect? *Research Policy, 33*(3), 425-441. https://doi.org/10.1016/j.respol.2003.09.004

Van Raan, A. F. J. (2005). Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods. *Scientometrics, 62*(1), 133-143. https://doi.org/10.1007/s11192-005-0008-6

Vieira, E., & Gomes, J. (2009). A comparison of Scopus and Web of Science for a typical university. *Scientometrics, 81*(2), 587-600. https://doi.org/10.1007/s11192-009-2178-0

Wenneras, C., & Wold, A. (1997). Nepotism and sexism in peer-review. *Nature, 387*, 341-343. https://doi.org/10.1038/387341a0

Ziman, J. (1996). Post-academic science: constructing knowledge with networks and norms. *Science Studies, 9*(1), 67-80.

Copyrights