



THE PROFESSIONAL REVIEW OF THE ITALIAN POLITICAL SCIENCE ASSOCIATION

ISSN 2420-8434

[\(https://italianpoliticalscience.com/\)](https://italianpoliticalscience.com/)

# Evaluating the Evaluation. The pros and cons of 'VQR' in social and political research

By Carla Monteleone, Stefania Panebianco, and Francesco Zucchini

[\(https://italianpoliticalscience.com/author/ipsmonteleonepanebiancozucchini/\)](https://italianpoliticalscience.com/author/ipsmonteleonepanebiancozucchini/), 15/06/2013

**ACKNOWLEDGING THE NEXUS** between science and economic development, and in the name of democratic control over the management of public resources, governments have progressively gained a role in mechanisms of knowledge production. In several countries, the United Kingdom, Australia, and France being the most well known cases, this has resulted in various evaluation exercises. All have generated wide debate within the scientific community on the most appropriate methods and criteria to be used. Italy, where the evaluation of research came a little later, is no exception. The debate has mostly taken place in the review *Il Mulino* and on the website Roars, with occasional articles appearing on the major national newspapers. Six key issues have dominated the discussion:

1) the definition of quality; 2) the drawbacks of impact indexes; 3) the informed peer review method; 4) the definition of research products; 5) the *ex post* adoption of evaluation criteria; 6) interdisciplinary comparison. In the following chapters we briefly introduce these topics with some comments.

## 1. The definition of quality

As for the definition of quality, the debate has focused on the inherent tension between quality as an objective fact and quality as a social construction. However, such a theoretical distinction is blurred when we pass from conceptual discussion to the empirical operationalization. In other terms, when we are interested in methods that are used to assess the 'quality'.

The most 'scientific', and seemingly objective, assessment procedures of quality are those based on bibliometric data. Since bibliometric data are numbers, you are unconsciously led to consider such procedures as objective. But these numbers are in some way created by the scientists themselves through the practice of quoting. Therefore these data are in fact inter-subjective evaluations. Moreover, there is a large variety of bibliometric data and different types convey slightly different information such as the different wording in a survey. They measure the impact on a certain scientific community of its members' products. Another distinction could be based upon the traditional tension in social research between qualitative and intensive methods and quantitative and extensive methods. Qualitative methods such as peer review are centered on scientific community evaluation as well. However, they measure the level of '**liking**', or acceptance. An author can be quoted because of his/her mistakes while on the contrary he/she is praised in a review only if his/her contribution is liked. On the one hand, **the impact** is inferred from a very large universe of cases, potentially the whole universe of scholars in a certain research field, but may convey ambivalent information. On the other, the level of '**liking**' inferred from a peer review is much less ambiguous, but it is usually

based upon the judgment of a very narrow set of referees that do not represent, necessarily, the prevalent opinion among experts with the same scientific credentials.

In fact, the impact indexes themselves contain some information about the 'liking' to the extent that these measures relate only to articles published in scientific journals where peer review method is adopted. Citations of these articles are considered simply because these articles are published. They are published because they are liked by a narrow set of experts.

## 2. The drawbacks of impact indexes

Other drawbacks of these measures have been the focus of discussion. First, the scientific quality is revealed over time. Impact measures can punish prematurely scientific products not yet sufficiently understood and appreciated for their value by the scientific community. In addition, the amount of citations obviously depends on the number of researchers who deal with a particular topic. A publication may have a considerable impact over another because the topic attracts more scholars than another topic and not for its intrinsic qualities. For example, in Political Science students of International Relations form a much wider community of scholars than students of Italian Politics and they are *a priori* likely to be quoted more often. Finally, the impact may ultimately reflect the extension of a network of scholars headed by powerful academics. In other words, the number of citations may reflect the level of subordination of scholars who cite, instead of indicating the degree of innovation, originality and explanatory power of the cited publication.

These difficulties should not be underestimated nor exaggerated. However, they can be mitigated. Some would suggest using the impact index of the journals in which the articles are published instead of the articles' impact, in order to minimize the problems connected with the different popularity of the topics and the local academic power. Nevertheless, the use of journal rankings is as controversial as the method of the informed peer review.

## 3. The informed peer review method

The third debate has centered around the use of the informed peer review method for Humanities and Social Sciences. This method combines traditional peer review with a classification of journals and it drew criticism from opposing sides of the debate, namely from both those who are against peer review and those who oppose bibliometrics. In this way the worst of two worlds is attained. Critics of the first type have pointed out that it is since Adam Smith that a warning circulates regarding the risk that peer review can be controlled by the most powerful academic groups. Therefore, it is important to know how referees are appointed, and, once the evaluation procedure is over, to have data about the referees, the number of research products that each of them have evaluated and the distribution of their evaluation. It has also been suggested that referees should know in advance that their evaluations will be revealed. It has also been proposed to start a dialogue within the scientific community on whether alternative methods now under discussion, for instance, *ex post* review, peer-to-peer review, etc. can be of any use to improve evaluation exercises in Social Sciences. According to this second type of criticism the experience in other countries seems to suggest that journal classification produces standardization and opportunistic behavior among researchers, and discourages cross-fertilization and interdisciplinary research. For example, in Australia the original classification into three groups was abolished because it produced distortions that were deemed too serious, and was substituted by a single list that only distinguishes whether a journal can be considered scientific or not.

Giving up classifying scientific journals, however, is a choice based on specific country-based assumptions. For instance, this could make sense if we estimated that the probability of ignoring an important contribution published in a not very diffused journal is greater than the probability that a referee, devoid of information on the journal status, could overstate or understate an article because of personal idiosyncrasies, incompetence or lack of time. The prevalence of one of these dangers depends on size, pluralism, expertise, resources of the panel of reviewers and duration of the evaluation assessment.

In the evaluation process in Political Science (VQR), the informed peer review method has been used only partially. The monographs and edited books in Italian political science are still an important part of scientific production. Almost all are in Italian and there is no reliable information to classify the Italian publishers according to the procedures that are used to select manuscripts to publish. Unfortunately, with perhaps one exception, there are no scientific editorial committees sufficiently broad and plural to guarantee an authentic *ex post* quality control of what is printed. So, contrary to what happens in journals, the nature of the container generally provides poor information as to the referees, and peer review has come to be uninformed.

## 4. The definition of research scientific products

An even more radical dispute concerned what should be considered as a scientific research product.

The debate has focused on whether it is enough to follow a standardized procedure, for instance peer review or inclusion in the ISI database, for a product to be considered scientific, or whether only by analyzing the content is it possible to tell whether a product is scientific or not. This debate has important implications in terms of costs, because the former position has lower costs than the latter. It also allows for a faster process, and limits the incidence of subjective elements. But, opponents claim, it has the disadvantage of inferring the content of the scientific product, for instance the article, from its container, or the journal. Because of the importance and sensitivity of the issue, the CUN (National University Council) has recently launched a public consultation process in view of defining what should be considered as scientific criteria and research products.

However, the difficulties of the theoretical discussion need not dramatically affect the effectiveness of the practices. Scientific journals are normally read only, or prevalently, by experts in a particular field of knowledge and they both publish articles whose primary purpose is related to the advancement of knowledge and have mechanisms which are as neutral as possible to evaluate the quality of the article with respect to the realization of its primary purpose. Other products do not. It is up to the scientist to associate the medium chosen to distribute his/her scientific work to the proper public.

## 5. The *ex post* adoption of evaluation criteria

The debate has also concentrated on the *ex post* adoption of criteria. The assessment exercise in Italy started without any prior indication. At the time of the articles' publication, publishing venues were not classified. At the least, they could be formally indifferent to the researchers. Undeniably, the introduction of a new evaluation system always involves some adjustment costs related to contemporary criteria retrospectively applied to previous behavior. However, at least during the VQR in Humanities and Social Sciences, the fundamental role played by peer review should have mitigated these costs. No product has been excluded from the evaluation on the basis of criteria unknown at the time in which the product has been submitted to the VQR. Both supporters and critics presumably agree that these evaluation exercises, rather than certifying the status quo, have a transformative effect, influencing how a scientific community will behave in the future. And the new criteria must be effectively adopted, not just announced, if they have to be credible for the future. This also raises the question of who establishes the criteria and upon which basis.

## 6. Interdisciplinary comparisons

Finally, critics point also to the fact that non bibliometric units (GEV) could significantly diverge in their classification methods, favoring opportunistic behavior. The risk is that Universities and departments with the highest concentration of scholars evaluated by 'stricter' GEVs may receive less funds, with long-term consequences on their possibility to grow.

In principle, the assessment should always be carried out only within the same disciplines. The publishing practices and assessment criteria vary greatly among different disciplines and make the same bibliometric indicators useless for such comparisons. It makes no sense to imagine that political scientists "contend" with lawyers, philosophers and natural scientists. Therefore, the distribution of funds between disciplines cannot be based on an evaluation exercise and is inevitably a political choice. Political scientists should not be scandalized but rather create a well heard and prestigious advocacy coalition with other social scientists to maintain, and possibly to increase, the proportion of funds diverted for the benefit of our research and studies.

On all of these issues, the debate is still open and is to be welcomed, because evaluation exercises are terribly complex and involve very sensitive issues. After all, if we study these phenomena we should know how to do it.

---

#### Share this:



(<https://italianpoliticalscience.com/2013/06/15/evaluating-the-evaluation-the-pros-and-cons-of-vqr-in-social-and-political-research/?share=twitter&nb=1>)



(<https://italianpoliticalscience.com/2013/06/15/evaluating-the-evaluation-the-pros-and-cons-of-vqr-in-social-and-political-research/?share=facebook&nb=1>)



Be the first to like this.

---

#### Related

<p>A new attempt at evaluating Italian research: the 2012-13 VQR (https://italianp... new-attempt-at-evaluating-italian-research-the-2012-13-vqr/) In "Home"</p>	<p>Evaluating the evaluators. A comment on the final report of the Group of evaluators of the area 14 - Political and Social Sciences (Call VQR 2004-2010) (https://italianp... the-evaluators-a-comment-on-the-final-report-of-the-group-of-evaluators-of-</p>	<p>An 'outside' perspective on the research assessment exercise (https://italianp.. outside-perspective-on-the-research-assessment-exercise/) In "Home"</p>
--	---	---