‘Science’ has always been evaluated. . . and will always be

Yves Gingras
Université du Québec à Montréal (UQAM), Canada

Abstract
Science has always been evaluated and will always be. The reasons justifying evaluations and the methods used to carry them have varied over time in relation to the many transformations of the sciences over the last fifty years. The uses of bibliometric methods are not limited to ‘evaluations’ of scientists and their institutions as they also provide a unique way to map global trends. One cannot stop evaluating science, but one can use the right tools at the right scale (individuals, institutions, countries) to better understand the dynamic of scientific change.

Keywords
bibliometrics, evaluation, scientometrics, university

Olof Hallonsten promotes the idea that we should ‘stop evaluating science’ (Hallonsten, 2021). Though he ‘self-evaluates’ his essay as being ‘provocative’, I think most readers will find the content rather classic and just reiterating, though usefully, well-known
analyses and complaints about the observed transformations of the scientific field since the 1980s.

I generally agree with the broad survey he presented, which is useful for introductory courses in Science and Technology Studies (STS). For instance, I often present the transformations he describes by contrasting two documents that define the two implicit ‘social contracts’ between science and society that existed before and after the 1980s. The first is the famous 1945 report prepared by the MIT engineer Vannevar Bush for the President of the United States, Harry Truman, and titled ‘Science, the endless frontier’ (Zachary, 1999). This document encapsulated the idea of an autonomous ‘scientific community’ or a ‘republic of science’ (Polanyi, 1962) that is best placed to decide on the uses to make of the money governments choose to invest in scientific research. Dominating the period 1945–1975, this vision lost its credibility during the 1960s and was replaced in the following decades by its exact opposite: science at the service of society, a program incarnated in the 1998 European Commission ‘vision of research and innovation policies for the 21st century’ titled: Society, the Endless Frontier. Notwithstanding the bizarre use of the metaphor of frontier that meant a lot to an American imagination but made no real sense in Europe, the choice of that precise title was meant as a clear message that the ‘republic of science’ had now been hijacked at the service of a new master, ‘society’, whatever that may mean precisely given the obvious sociological fact that it is composed of many conflicting groups acting in a diverse array of fields and having quite different interests and logics (Bourdieu, 1975, 2018). This decline in autonomy of the scientific community also became evident in the new tools used to evaluate the scientists and institutions that deserved promotions or money (Gingras and Khelfaoui, 2021). It is on these ‘problematic aspects’ of the current science policy regime and on ‘the ubiquity of quantitative evaluation practices’ that Hallonsten rightly focuses his attention.

In order not to repeat his critiques, with which most academics will easily agree, I will focus my brief comments on what I think is missing in such global and very general discourses on evaluation.

The functional role of evaluation and control

Let us begin by recalling that the basic functional reason why evaluation of research (publications and grants) has always existed and simply cannot disappear is that it serves as a necessary selection mechanism in a finite system with finite resources confronted to a potentially infinite offer. These resources can be money, personnel, instruments or time. Hence, even the initial utopia of a non-controlled Wikipedia or arXiv, showed that control and evaluation were needed in order to make those projects really work and not be submerged with absurd affirmations of all sorts with no ‘supreme court’ to eventually stop the endless fights and controversies. Likewise, preprint servers like bioRxiv and medRxiv do not automatically accept all submissions. And even among the peer-reviewed papers published in scientific journals, their number is so large that no scientist can even read all those in his/her own limited specialty and must thus find a criteria of choice based on a form of evaluation: suggestions made by trusted colleagues, the prestige of the journals, etc. So, evaluation has always existed and serves many functions: it censures discourses considered dangerous by states (the imprimatur), by churches (The
also before looking at the new forms of evaluation that have emerged over the last thirty years, I think it is useful to unfold a bit the vague notion of ‘science’. One usually says that scientists do ‘science’ in that they have a scientific practice aiming at producing new knowledge. But to be considered knowledge and not mere belief, any such proposition must be evaluated by the peers of the same scientific specialty. But one can also say that institutions (departments, universities, research laboratories) conduct scientific research and it can be meaningful to track these practices at the aggregate level of these institutions. Quantitative indicators then provide a map of the level of activity and thus provide also a kind of evaluation of what is going on in these large organizations. Finally, one can also say that countries are engaged in science and here again compiling bibliometric data can be helpful. They make visible for instance that a country like China has greatly increased its scientific activity over the last twenty years, while other countries like Russia, for instance, have seen their research system much affected by the dismemberment of the USSR (Kirchik et al., 2012). So, while aggregate bibliometric data can help ‘evaluate’ the level of science between countries, no individual scientist can provide a valid global view, as he or she is always embedded in a local niche and thus subject to the bias of ecological fallacy – which tend to take a local sample as representative of the global scale.

Finally, despite the dominant discourses on ‘globalization’, one must acknowledge that the effects of the ideology of New Public Management – that has tried since the 1990s to impose itself on universities – have not been the same in all countries. For instance, the impact of the publication since 2003 of the so-called ‘Shanghai Ranking’ of the ‘best’ universities of the world has been much more important in Europe than in North America. In France for instance, it has been used as an opportunistic weapon by the Sarkozy government to justify in 2007 a large university reform (Gingras, 2008; Barats et al., 2018; Harari-Kermadec, 2019). In Australia and in some European countries, bibliometric indicators have even been incorporated into funding formula, thus creating now well-known perverse effects (Gingras, 2016). The uses of such quantitative indicators also differ much between disciplines, with biomedical sciences and economics being more adept to such simplistic – and in fact invalid – indicators, like the ‘impact factors’ and the ‘h-index’, than mathematics or the social sciences and humanities. And far from being promoted only by bureaucrats and managers, the use of flawed indicators as measures of ‘quality’ is also actively put forward by scientists in their struggle with their opponents and competitors.

The right indicator at the right scale

What the general critiques of the uses of quantitative indicators of research too often miss is that those indicators do not have to be identified with ‘evaluations’ and even less with ‘rankings’. For evaluations are not synonymous with or equivalent to rankings, as
only the latter is a public announcement of a list of ‘top’ institutions. Such rankings are attractive because of their extreme simplicity: a single number! Such ‘league tables’ transform multidimensional institutions into one-dimensional ones, along the lines analyzed by Herbert Marcuse in his book *One-Dimensional Man*, more than fifty years ago (Marcuse, 1964). But, as noted above, bibliometric data can be useful to map the dynamic of science and compare countries and institutions. And again, comparing is not the same as evaluating if the latter notion implies sanctioning. Comparing helps understand why institutions may differ. Also, an indicator that is valid at an aggregate level could be misleading at the individual level. There is thus no need to invoke the vague notion of ‘complexity’ that ‘quantitative performance evaluation’ could not capture. Like a thermometer captures temperature but not humidity, for which one needs a hygrometer, one simply has to understand the limits of the instrument used in order to correctly interpret the information it provides.

We can only agree with Hallonsten that ‘excellence’ is an empty rhetoric aimed at convincing governments that the money is well spent. In practice, the notion is tautological as all organizations say they choose only the ‘excellent’ projects and people. Some granting agencies accept 10 or 15% of applicants while others can accept 35% or even 70%, and they all say they choose ‘excellent’ projects (Gingras, 2019). Similarly, the notion of ‘evidence-based’ policy is also problematic for, here again, one tends to accept the ‘evidence’ only when it is in accord with one chosen policy and attack the methodology of any ‘evidence’ when one does not like the results and decisions (Gingras, 2017).

We also agree that ‘the burden of proof should be shifted over to those who claim that science is insufficiently productive’ (Hallonsten, 2021: 9), but think that this does not have to be linked with the question of the evaluation of science. For one can simply insist and recall that the basic aim of science is not to be ‘productive’ – whatever the exact meaning of that term, which usually suggests only ‘economic productivity’ – but to provide validated knowledge on the natural and social worlds. That such knowledge may sometimes also be transformed into economic or social gains is a different matter. For one can certainly argue that each kind of social organization has a specific role and function and that transforming universities into firms or ‘innovation hubs’ would only deflect them from their basic mission. Likewise, one should not ask private firms to necessarily engage in basic research and provide open and free access to their knowledge, as that is not their main function. In short: the problem is not ‘evaluation’ per se but using measuring instruments that are adequate to the mission of the institution being analyzed and, in this sense, ‘evaluated’.

**Funding**

The author received no financial support for the research, authorship, and/or publication of this article.

**References**

Barats C, Bouchard J, Haakenstaaad A (eds) (2018) *Faire et dire l’évaluation. L’enseignement supérieur conquis par la performance*. Paris: Presses des Mines.

Biagioli M (2002) From book censorship to academic peer review. *Emergences* 12(1): 11–45.
Bourdieu P (1975) The specificity of the scientific field and the social conditions for the progress of reason. Social Science Information 14(6): 19–47.
Bourdieu P (2018) Classification Struggles. General Sociology, Volume 1 (1981–1982). London: Polity Press.
Gingras Y (2008) La fièvre de l’évaluation de la recherche. Du mauvais usage de faux indicateurs. Montréal: Note de recherche du CIRST, 2008-05.
Gingras Y (2016) Bibliometrics and Research Evaluation. Uses and Abuses. Cambridge: The MIT Press.
Gingras Y (2017) Evidence-based policies for all – Except for science? University World News, 8 September. Available at: https://www.universityworldnews.com/post.php?story=20170906110636466
Gingras Y (2019) Ordonner les universités : une affaire classée ? Natures Sciences Sociétés 27(3): 359–361.
Gingras Y, Khelfaoui M (2021) L’effet SIGAPS : la recherche médicale française sous l’emprise de l’évaluation comptable. Zisel 8: 145–174.
Hallonsten O (2021) Stop evaluating science: A historical-sociological argument. Social Science Information 60(1): 7–26.
Harari-Kermadec H (2019) Le Classement de Shanghai. L’université marchandisée. Paris: Bord de l’eau.
Kirchik O, Gingras Y, Larivière V (2012) Changes in publication languages and citation practices and their effect on scientific impact on Russian science (1993–2010). Journal of the American Society for Information Science and Technology 63(7): 1411–1419.
Marcuse H (1964) One-Dimensional Man. Boston: Beacon Press.
Polanyi M (1962) The republic of science: Its political and economic theory. Minerva 1(1): 54–73.
Sugimoto CR, Larivière V (2018) Measuring Research. What Everyone Needs to Know. Oxford: Oxford University Press.
Zachary GP (1999) Endless Frontier. Vannevar Bush, Engineer of the American Century. Cambridge: The MIT press.

**Author biography**

Yves Gingras is professor of history and sociology of science at the Université du Québec, Montréal, and scientific director of the Observatoire des sciences et des technologies (OST). His most recent books are Sociologie des sciences (PUF, 2020, 3rd edition) and Histoire des sciences (PUF, 2021, 2nd edition).