

22nd Ethical Forum of the University Foundation
Towards a revolution in the assessment of academics?

CAN WE ESCAPE THE INDEX TRAP ?

Personal conclusions
by Philippe Van Parijs
UCLouvain & KU Leuven

The 2024 edition of the Ethical Forum was coordinated by Professor Jean-Paul Lambert, honorary rector of the Université Saint Louis and chair of the Académie de recherche et d'enseignement supérieur of the Fédération Wallonie-Bruxelles. Keynote background lectures were delivered by Professor Véronique Halloin, general secretary of the Fonds national de la recherche scientifique, and by Professor Bert Overlaet, former director Human Resources at KU Leuven and author of the position paper of the League of European Research Universities on « A pathway to multidimensional academic careers ».

These lectures were followed by a panel moderated by Professor Helder De Schutter, chair of the Higher Institute for Philosophy of the KU Leuven, with the participation of Professor Françoise Smets, rector of the UCLouvain, Professor Isabel Van Driessche, dean of the Faculty of Sciences of UGent, Professor Lieve Van Hoof, professor at UGent and former chair of the Jonge Academie, and Dr Thérèse Zhang, deputy director for higher education policy at the European University Association.

All the speakers' presentations and various documents they referred to are available on the [site](#) of the Ethical Forum of the University Foundation. I am most grateful to all of them for much information and many insights. The text below is not meant to synthesize their contributions but, more modestly, to formulate some reflections stimulated by what I learned from them.

The quantitative blind alley

“Decisions to appoint and promote academics rely excessively on quantitative indicators of research performance.” This is a conviction that is widely shared among academics, both among older ones, who can remember a less index-obsessed era, and among younger ones, who hate to see their fate sealed by a couple of indicators. It is also a conviction that made its way to the top of their universities, to university associations, and to European policy makers.

Thus, Bert Overlaet started his presentation with a quote from Marc Schiltz, president of Science Europe : “‘Publish or perish’ and metrics have led us into a blind alley.” He went on to recall the core recommendations of the Agreement on Reforming Research Assessment (ARRA) of 20 July 2022. Two of them read as follows: “Base research assessment primarily on qualitative evaluation for which peer review is central, supported by responsible use of quantitative indicators” and “Abandon inappropriate uses in research assessment of journal- and publication-based metrics, in particular inappropriate uses of Journal Impact Factor and H-index”.

The Coalition for Advancing Research Assessment (CoARA) consists of over 700 universities and other academic institutions that signed the ARRA agreement. Its Working Group on Reforming Academic Career Assessment, co-chaired by the European University Association and the Young Academy of Europe, insists that the assessment of academics “should adequately reflect the different tasks, functions and roles academics fulfil over the course of their career”.

An evolution rather than a revolution

Such declarations suggest that a revolution is in the making, a revolution that would enable European universities to escape from the “blind alley” of quantitative indices of research performance as the key determinant of academic careers. However, speaking about an imminent revolution can be misleading. Bert Overlaet, Véronique Halloin, Françoise Smets and Isabel Van Driessche all stressed that quantitative indices have never been the sole way of assessing research performance in our academic institutions. And evaluation criteria based on other achievements than research performance have always been present in our universities.

However, none of our speakers denied that sometimes quantitative indices had been playing too prominent a role, and they therefore all welcome and support CoARA’s initiative. Rather than a revolution that would shatter the practices of their institutions, one should speak of an evolution to which they are happy to contribute by making recruitment and promotion procedures in their respective institutions less index-dependent and, in universities, less exclusively research-focused. As Lieve Van Hoof illustrated, however, this evolution sometimes operates through micro-revolutions experienced as so many liberations by their beneficiaries.

Despite this unanimous agreement on the importance of granting appropriate weight to the university’s teaching and service missions and of submitting research to qualitative peer review, there seemed to be a broad consensus to the effect that research performance should remain the primary consideration at the appointment stage — and obviously the only one for research agencies like the FNRS — but also that quantitative indices of research performance cannot be dispensed with entirely. This suggests that determining the extent to which quantitatively assessed research performance needs to be taken into account will remain a permanent balancing act. Moreover, various remarks suggested that we should expect a structural pressure to fall back on handy indices. In this area as in any other, developing a realistic vision of a desirable future requires us to understand the root causes of the undesirable situation from which we are trying to escape.

The podium

What is it that led to the prevalence, now judged excessive, of quantitative indices? Two distinct developments, it seems to me, both quite positive in themselves: the international mobility of students, which created what I shall call the podium trap, and the phasing out of inbreeding, which created what I shall call the clock trap.

The growing international mobility of students triggered the development and popularization of university rankings (to which the University Foundation devoted its 2009 Ethical Forum). The process started in 2003, with the launch, by Shanghai Jiao Tong University, of the Academic Ranking of World Universities (ARWU). It was largely motivated by Chinese institutions’ wish to know where it would be most fruitful to send their best students,

especially in the fields of science and technology. The practice soon spread to all academic disciplines, and competing Western rankings emerged, starting in 2004 with the Quacquarelli-Symonds-Times-Higher-Education ranking, in an effort to better capture the specificities of all of subjects and all institutions.

As measurement of the quality of universities, the methodology is very sloppy and unavoidably unfair — which poorly ranked institutions are particularly keen to stress. Yet, however sloppy and unfair, the rankings cannot be ignored. They affect the attractiveness of a university for international students, academics and partnerships. As they tend to be visibly echoed in local media, they also impact the self-esteem and morale of institutions, their leaders and their members. And they are even sometimes used by national agencies for the allocation of their funds.

Whether directly or indirectly, research performance dominates the rankings, from the number of Nobel laureates to the reputation among foreign colleagues, hardly ever based on anything other than the notoriety gained through scientific publications. The number of citations overall or per capita also often features explicitly as a weighty component. In this light, can one really expect universities to resist the pressure to try to recruit potential “high performers”, i.e. researchers likely to help the recruiting institution move higher in the rankings? And can one expect them to refrain from adopting promotion rules that would help retain those “high performers”?

Of course, universities realize that high performance is a group achievement, with invisible data curators or software managers as important as the most often cited scholars. But the powerful, understandable and diversely motivated desire to climb on the podium erected by unignorable rankings will force each university — and thereby also, in mutually reinforcing fashion, all those with which it competes — to give a high priority, in the assessment of its academics, to quantitative indices whose impact on the rankings is beyond doubt.

The clock

This pressure stemming from the rankings is further strengthened by a second development: the phasing out of inbreeding, i.e. the universities’ increasing appointment of academic staff from outside, and often far outside, their ranks. Like international student mobility, this is a development that is, as such, to be welcomed. It enables universities to find better matches for the vacant positions, to diversify the background of their members and to reduce the risk of nepotism. But like international student mobility, it has a perverse consequence.

Choosing among a few people whose student grades one has no difficulty interpreting, whose dissertation one has evaluated, whose talks one has attended, whose papers one has read, whose personality one knows, is neither complicated nor time-consuming. Choosing in similarly informed fashion among dozens, sometimes hundreds of applicants from all over the world is quite another matter.

For example, CoARA pleads for “qualitative judgement” and for the now common practice, referred to repeatedly during the Forum, of asking candidates to send a “narrative CV” (now possibly prepared with the kind assistance of AI) and what they regard as their five best publications. But how many, among the members of the selection or promotion committees, read more than the CV and the abstracts? How many, among the members of the selection or promotion committees, are familiar enough with the candidates’ research domain to be able to assess how innovative these publications are or how well they meet the standards of the relevant scientific communities?

In some cases, for example when the UGent allows for a fast-track to promotion as an exception to its seniority-based flat-career rule, the committee circumvents this problem by using the criterion of having obtained an ERC grant. This amounts, somewhat uncomfortably, to letting an external body decide who gets promoted in one's own institution. And above all, it only displaces the problem. For the ERC committees are also composed of a set of people who cannot be competent in all the subfields of the projects they have to assess. Moreover, the committee members must be people willing to spend several days reading the files and attending the meetings. How likely are they to be among the best scientists in the field? And if and when they are, what a high opportunity cost in terms of managing their teams, supervising their researchers, teaching their students, doing their own research!

This illustrates the fundamental dilemma of qualitative assessment after the end of inbreeding. If the evaluation is done seriously by competent academics who take the trouble of reading thoroughly the material provided by all the candidates and of acquainting themselves sufficiently with the relevant domain to be able to assess the quality and originality of that material, then the value of the time spent on it is prohibitively high. While they do all that work conscientiously, the clock is ticking, and the committee members' many other academics duties remain unattended. But if one does not devote to this task the huge amount of valuable time a good qualitative evaluation requires, its quality is bound to be worse than one using quantitative indices.

Quantitative indices as outsourcing of qualitative assessment

Why ? Basically because using such indices amounts to outsourcing the time-consuming qualitative assessment to people who can do it better than the average member of a selection or promotion committee and in fact have already done it: the many journal referees, who accept or reject articles and thereby determine the number of peer-reviewed articles on a candidate's CV, and the even larger number of scholars in the relevant field who find or do not find articles sufficiently interesting to cite them. This path leads straight to the infamous H-index.

As recounted by Véronique Halloin, the H-index made its appearance in 2005 and was systematically used by Web of Science from 2007 onwards and by Google Scholar from 2011 onwards. A researcher's H-index is the number H of her/his scientific publications that have been cited at least H times in other scientific publications. It offers a simple combination of a measure of "productivity" (the number of publications) and a measure of "quality" (the number of citations). Other indices can bring in the quality of the journals in which the articles are published, as measured by their impact factor (the average number of citations per article published in them) to weigh the researcher's articles or even the articles in which it is cited.

Whether or not the impact factor enters explicitly in the index, evaluation practices using such indices have the very unwelcome perverse effects of subjecting researchers and their institutions to the greed of the for-profit publishers that own high-impact-factor journals. Moreover, they obviously miss a lot about the qualities of a researcher, and even more about the qualities of an academic in all dimensions. It is this widespread conviction that motivated the CoARA initiative and other initiatives in the same direction. In theory, an in-depth qualitative assessment of all candidates based on a careful and competent reading of their best publications followed by a live interview by top experts in his field without personal connection to them provides an extremely attractive alternative to the use of quantitative

indices. In practice, however, for the reasons sketched above, actual “qualitative” assessment is generally very remote from this ideal.

As a consequence, the use of quantitative indices such as the H-Index is clearly, if not a universally superior alternative, at least an indispensable complement for filtering out totally unsuitable candidates, for drawing up a short list and perhaps even for singling out one candidate whom one would need to adduce very strong reasons not to appoint. Time is scarce. In the post-inbreeding era, the prior evaluation work performed by many competent colleagues and summarized by quantitative indicator cannot be wasted if one is to achieve satisfactory assessments at a reasonable opportunity cost in academic time.

How not to slip back

Because of the double pressure of rankings (“the podium”) and time scarcity (“the clock”), there is no prospect of getting rid entirely of quantitative indices. But can we protect ourselves against slipping again towards excessive reliance on them at the expense of missing scientific qualities that they do not capture and above all at the expense of giving appropriate weight to the academics’ and academic institutions’ other missions? The pressure of time scarcity could conceivably be relieved somewhat by using AI to digest bulky files more quickly and to compare them more fairly. But the opportunity cost of conscientious evaluation will remain considerable. If loosening the grip of the clock is unpromising, can more be expected from reshaping the podium?

As explained by Thérèse Zhang after the Forum, providing better transparency for the fulfilment of all the missions of our universities was one of the objectives of the U-Multirank initiative, launched in 2011 and funded by the European Commission. U-Multirank provided grades ranging from A to E to universities, on the basis of self-reported data along five dimensions (teaching, research, knowledge transfer, international orientation and regional commitment). It did not offer composite scores that might have provided the basis for widely advertised rankings in the same vein as ARWU, Q&S and THE. U-Multirank raised interest among education policy makers, but remained under-noticed by a broader public. In 2023, the European Commission decided to merge it in the European Higher Education Observatory, which aims to collect a wide range of data about higher education in the European Union and beyond, without any ambition to provide alternative rankings. (See the European University Association’s [*Key considerations for the use of rankings by higher education institutions*](#), 2023).

Consequently, the prospects for reshaping the podium are not very brilliant either. What is left? A coalition of the willing such as CoARA endeavouring to transform the international academic ecosystem, to use Bert Overlaet’s formula. Our universities know that, if they are to do well in the rankings, they need to attract and retain many quantitatively certified “high performers”. But they also know that the adequate fulfilment of the whole range of their missions requires them to recruit and reward a wide diversity of other profiles. If each of them does what is best for the whole spectrum of their missions, there will be less pressure to pinch top performers from one another, less perverse ranking-driven war on talent to the detriment of our universities’ overall quality.

True, if non-EU universities — especially British and American ones — do not join the effort, the temptation to defect in order to halt the scientific brain drain to Anglo-Saxon countries will be high. But it will be less than in the absence of a coalition. And apart from returning to the pre-mobility general-inbreeding era, there is no alternative.