Evaluation and Merit-Based Increase in Academia: A Case Study in the First Person

Christine Musselin

Valuation Studies 8(2) 2021: 73-88

Symposium contribution

Evaluation and Merit-Based Increase in Academia:A Case Study in the First Person

Christine Musselin

Abstract
This article provides a reflexive account of the process of defining and implementing a mechanism to evaluate a group of academics in a French higher education institution. The situation is a rather unusual case for France, as the assessed academics are not civil servants but are employed by their university and this evaluation leads to merit-based salary increases. To improve and implement this strategy was one of the author's tasks, when she was vice-president for research at the institution in this case. The article looks at this experience retrospectively, emphasizing three issues of particular relevance in the context of discussions about valuation studies and management proposed in this symposium: (1) the decision to distinguish between different types of profiles and thus categorize, or to apply the same criteria to all; (2) the concrete forms of commensuration to be developed in order to be able to evaluate and rank individuals from different disciplines; (3) the quantification of qualitative appreciation, i.e. their transformation into merit-based salary increases.

Keywords: evaluation, commensuration, quantification, merit-based processes, academics

Christine Musselin *is a Professor at the Centre de sociologie des organisations, Sciences Po, CNRS.*

https://doi.org/10.3384/VS.2001-5992.2021.8.2.73-88
Hosted by Linköping University Electronic press
http://valuationstudies.liu.se

Introduction

In this article, I build on my own experience (between 2013 and 2018) as vice-president for research at the institution where I work, Sciences Po. It is a reflexive analysis of an experience of 'valuation in practice'.

In France, the position of vice-president is held by academics who generally have previously been a research group leader, or department chair for a number of deans, but are not trained in leadership or management; they generally go back and work in their previous post after their mandate as administrator ends. This applies to me. The difference is that my academic training was in the sociology of organizations and my main field of study is higher education institutions and academic labour markets.

In particular, I have led a study on academic recruitment, where we inspected two disciplines (history and mathematic) and three countries (France, Germany, United States), and paid attention to the construction of judgement in hiring committees, as well as to the setting of the price (i.e. the mix of salary, starting funds and personal benefits) of the recruited academic (Musselin 2009 [2005]). In this study, I also collected information about the acquisition of tenure, the promotion processes to associate and full professorship, as well as on the yearly evaluation process and merit-based salary increases happening in US institutions where I studied private not-for profit universities. I had also led a study on access to professorships for female *maîtres de conférences* in France and a collaborative project funded by the ANR (French national research council) on the trajectories of French academics in physics, history and management sciences who had a first permanent position in the mid-1970s, 1980s, 1990s and 2000s, in order to identify what had changed in academic trajectories in terms of access and development (Musselin et al. 2015). Not only have I had personal experience as a member of hiring committees and diverse evaluation bodies, but I am also a researcher who has tried to understand valuation practices in different academic labour markets. In this article, I have tried to use the knowledge and concepts from my field of research to reflect on a situation in which I worked as a practitioner. Besides my own work, I rely in this on studies of valuation on commensuration and rankings, in particular by Espeland and Sauder (2007, 2016; see also the interview in this issue, Ossandón et al. 2021) and on the work on academic evaluations by Lamont (2009). The evaluation committee I had to manage was closer to the evaluation panels assessing research projects that Lamont studied than to the hiring committees in my own research as it is multidisciplinary and has to deal with different scientific registers.

The situation: assessing academics

Of the many valuation situations I encountered as vice-president of Sciences-Po, the most challenging was the evaluation of the academics directly employed by my institution.

Along with the civil service positions of academics employed either by the CNRS (National Center for Scientific Research) or by the Ministry for Higher Education and Research (the university professors), which can traditionally be found in French universities, the institution where I work employs about 70 academics hired with long-term private contracts. Academics in this position have their own career development and promotion schemes, even if, in order to maintain as much proximity as possible in the treatment of all these different populations, we try to keep the latter close to the regulations applied to civil servants.

In 2008, Sciences Po decided to organize recurrent evaluation for this group of academics[1] and to base the allocation of merit-based salary increases on the results of this evaluation, which deals with four different domains: scientific production, teaching, institutional and discipline-based involvement, and impact. It took place every two years (now it is every three), and each academic has to write a report on his/her activity during the relevant years. The result of the evaluation is transformed into grades for each of the evaluated domains and different coefficients are applied to these grades (more for research than for teaching, more for teaching than for institutional and discipline-based involvement, and more for institutional and discipline-based involvement than for impact). The final grade is based on a ranking on which the allocation of merit-based salary increases (from 0 to 10 per cent max) will be decided, coming on top of the annual basic salary increase allocated to every academic and non-academic employee who is not a public servant. The amount dedicated to merit-based salary increases was set before the process began and could not be increased.

When I took over the position of vice-president, two evaluation processes of locally contracted academics had already been running under my predecessor's aegis and I had to organize a third one immediately after my entry. I remember I looked at the procedures already in place and identified potential problems but did not have time to negotiate any transformation at that point. The complicated election of a new president after the sudden death of his predecessor had already delayed the setting of the evaluation committee for six months and it was not possible to delay it any longer.

---

[1] Civil service academics have their own evaluation processes. They are run at the national level for researchers of the CNRS and partly by a national body for university professors.

Nevertheless, organizing this evaluation with the existing rules was a fruitful experience as it gave me the opportunity to observe some of the structural weaknesses of the process. Three in particular struck me. The first two were related to the composition of the committee: formal aspects do matter. First, the presence of directors of the different labs (groups of researchers, equivalent to departments elsewhere) on the committee, which encouraged them to defend their own staff and try to persuade everyone that they only had exceptional scholars. This of course is linked to the competition among them for internal resources and the opportunity the committee gave them to campaign in favour of their own lab. But is also related to the fact they knew that any lack of support for one of their colleagues would immediately be known and diffused: they therefore preferred entertaining social peace and being nice to everyone. The second weakness came from the very low number of external reviewers and the room it left for internal games. The third was of a different nature. I was concerned by the poverty of many of the reviews prepared by members of the committee and by the lack of consensus on what should be taken into account for each of the four domains. Part of it came from the fact that the committee was rather ad-hoc and made up of individuals coming from different disciplines and not used to making joint decisions – which complicates the development of routines – and had to learn to work together. But I attributed it also to another rather classical issue in quality assessment (Musselin and Paradeise 2005 [2002]; Beckert and Musselin 2013): uncertainty is not only about evaluating the level of quality itself but also about identifying and agreeing upon the dimensions that define the quality of an activity and which have to be taken into account in the evaluation process.

This convinced me that the evaluation process should evolve: in terms of the composition of the committee and in setting rules about conflicts of interest and how to deal with them. I built on what I had learnt from my research on hiring committees and on my participation in other evaluation bodies to change the composition and to suggest rules of conduct; but this is not the topic of this paper. I will therefore rather concentrate on the aspects that are more directly connected to valuation practices and focus on categories, commensuration, evaluation and valuation, and quantification.

Categorizing

Two issues arose in terms of categorization. The first is linked to the fact that this group of academics was not any longer homogeneous. These 'home-made' academics were created in the 1950s and built on the model of the CNRS researcher: they could of course teach but did not do much of that until the 2000s, and then always on a voluntary basis, as they only had research duties. But in 2009, it was decided that

the newly recruited academics would have both teaching and research duties and would be called professors.[2] Those who had been recruited as researchers were offered the opportunity to 'converge' (an exact colloquial translation of the term we used) and to join the new status of professor with a rather substantial increase in their salary, but also with regular teaching duties. The conversion was nevertheless not automatic and a specific procedure was deployed in order to examine the applications: some of them had been refused either because the applicant had too little teaching experience or because his or her research records were not considered as sufficiently satisfactory. I must add that the first three evaluation rounds, including the one I led when I just arrived, only concerned the 'researcher group' because the procedure for organizing the evaluation of the professors was still to be written. I therefore engaged in a reform process and set up a working group of academics of different status (including some CNRS researchers and university professors) and from different disciplines in order to extend the evaluation to the professors, both those recently recruited and researchers who had become professors.

The first decision to make in terms of categorizing, was thus to deal with the recognition (or not) of a distinction between professors and researchers. Should we have only one evaluation scheme and one merit-based allocation framework for all? This was important because it was a way of acknowledging reality but also a way to bring together and therefore bring closer the two groups which might have been considered as very different. It was clear that not distinguishing between the two groups was a way of sending a signal to the researcher group that they should aim for the new status, and that otherwise they would always be exposed to receiving an 'unsatisfactory' grade regarding teaching, thus getting less chance to be at the top of the pile for the merit-based increases. But, as mentioned above, some of the applications to shift status were refused and it could be expected that some would never be accepted: there was a kind of contradiction in both inciting to 'converge' while restricting access to the new status. Another point was that a number of the researchers were giving some classes even if they were not obliged to, and sometimes as much as (or more than) those with teaching duties, and nobody wanted to discourage them from doing so.

Thus, categorizing is also compromising between different logics and constraints. This led the working group to opt for maintaining a distinction between two categories that do not exactly follow the researcher/professor divide. The first category regrouped researchers as those having given fewer than two classes per year in the last three years – the so-called pure researchers – and the second included all

---

[2] They could be assistant, associate or full professors, as a tenure track system was introduced at the same time for this category of academics.

professors and researchers having taught more than two classes per year over the last three years. By recognizing that some of the researchers are also teachers, although they do not teach as much as professors, or do teach as much but do not want to become professors,[3] the institution recognized their involvement in teaching and encouraged them to continue it – while there are very high teaching needs – by including this activity in their evaluation and by considering them as 'equivalent' to professors.

The second decision in terms of categorization concerned the identification of the domains to be evaluated. We wanted to make clear what is expected of the reviewers as well as for those under review. This formalization is a trend that can be observed in many evaluation bodies nowadays: they depart from more impressionist views and develop templates that applicants on the one hand and reviewers on the other have to fulfil. Efforts are thus led to identifying what is expected and what is not. Those expectations of course reflect what is important, what is deemed worthy by the evaluation organizers.

As mentioned above, the first two evaluation committees and the documents regulating them suggested that four types of activities should be assessed, i.e. are expected to be achieved: research production, teaching, institutional and discipline-based involvement, and impact. There has been no discussion within the group about changing the four domains, but effort has been made to better define which activities belong to which, i.e. which elements will be assessed or what are the components of quality for each domain. Reciprocally, of course, this led to ignoring some other elements or even to deliberately excluding them: for instance, nobody claimed that book reviews should be included in the publication records, or that being a member of a professional association was a relevant indicator of involvement in a discipline. This phase, however, was not only about selecting indicators and leaving others behind; it also meant attributing activities to domains. For instance, it was decided that being an editor of a journal relates to the third domain rather than to the first one, while obtaining grants belongs to the first rather than to the third. This resulted from discussions in the working group. Then, assessing each of the elements that constitute quality and integrating them into a synthetic evaluation is another issue that can only happen in action, or in interaction within the committee.

---

[3] This might seem curious as what they get in terms of revenue for their teaching was clearly less interesting than what they could get in salary increase by becoming professors, but with the new status they became obliged to respect the teaching duties every year and had less freedom in terms of choosing what to teach, as professors have at least one class at the bachelor level and one in the regional campus. Their freedom to organize seems to them more valuable than the increase in revenue they could achieve

It should be said that all this work of formal definition, of what will be taken into account and evaluated, is important for both the reviewers and for those will be evaluated. The latter know what is evaluated and can develop their activity report accordingly, while reviewers are aware of what they should look for when reviewing. The choice of items is not technical but axiological and practitioners should be aware of the implicit values embedded in what seems to be a neutral instrument. This objectification is always biased, in the sense that it implicitly favours some profiles over others. In the activity report's template, for instance, we ask for a description of research projects overseen in the last three years as well as publications over the same period. This kind of demand is not favourable to those who work on long-term fieldwork, or to those who prioritize books over papers. Applying for grants can also be detrimental to those who do not need a huge amount of resource for their research. Illness or pregnancy is also difficult to take into account, because such templates tend to be biased toward linear productivity.

Categorization in the four domains led to the construction of evaluation frameworks fixing the weight of each dimension in the final assessment. The constitution of two categories of evaluated scholars already discussed led to the construction of two evaluation frameworks. Researchers with none or few teaching activities are evaluated only for research, institutional and discipline-based involvement, and impact, while professors and teaching researchers are evaluated for the same activities plus teaching. The next step was to then set the parameters to be applied to each domain. The question behind this was of course what priority should be given to some activities over others. Even if the four dimensions for the new status and the three for the 'pure researchers' are all things that are expected to be achieved by each and every one, none is worth the same weight in the evaluation.

I expected that the problem of how to weight the different dimensions would provoke much discussion and some negotiation between different representations of the desired profiles – i.e. what each member of the working group considers as an 'ideal' repartition of the expected activities. But this was not the case. Actually, nobody claimed that research was not the more important task or that people should not first be rewarded for their academic production. Weighing research as half of the evaluation was quickly agreed, probably because 50 per cent worked as a magic number. As a kind of counterpoint, nobody pleaded for a higher coefficient than 5 per cent for impact. This clearly reflected a shared representation among academics in the working group about their role – they first of all find their justification in research and fundamental research – and the kind of institution they want to be associated to – a research university caring about impact but first rewarding scientific results per se.

Valuation was also therefore a way to symbolically defend a definition of one's job; some impact but first of all research.

The decision that 'pure researchers' should not be expected to be more engaged in impact and institutional and discipline-based involvement furthermore reflects the idea that the time they save by not teaching should not be devoted to these activities, but that higher research requirements will be expected from them. This at the same time increased expectations about their research achievement and, in the end, the effects of poor research results counts for 80 per cent of their evaluation. Some members of the group explicitly expressed this consideration and supported this expectation.

Commensuration: the crucial role of a committee chair under control

Categorizing the types of scholars and weighing the criteria to apply to differently evaluated domains informed what is expected but not how to evaluate and how to come to a judgement on each domain. In order to come to this, each individual activity report was sent to two reviewers, one internal, one from a different university. They would have to return their reviews before the career committee (which is made up of all reviewers, half of them external and half from the institution) could meet. They also had access to all activity reports. The committee was multidisciplinary and there were representatives from the five main disciplines covered in Sciences Po: law, history, sociology, political science, and economics. The attribution of reviewers generally closely respected the discipline of those whose work was assessed, and if that was not possible, at least one of the two reviewers would belong to the main discipline and the other would be from a different area.

For each review, on each domain, we asked for a grade and a written assessment that clearly documented the chosen grade. Four possibilities were given: A for outstanding, B for excellent, C for satisfactory and D for unsatisfactory. The reviewers met for about two days and each case was discussed in succession.

Because of the tight schedule,[4] it was not possible to open up the floor to each reviewer, followed by general discussion. We decided to keep the same process that had been used previously – where the vice-president for research chairs the committee and presents, for each researcher a brief summary of the two reports, domain after domain. So, for instance, I summarized the reviewers' written arguments on Mrs Clare's scientific production, and said that one awarded a B and the other an A, and that I would propose to consider Mrs Clare's work as 'excellent' (i.e. to give her a B, as both reviewers in their comments outlined the high quality of the work achieved but none of them made

[4] It was very difficult to ask external reviewers to stay longer than two days, especially, because they were not paid remunerated for this task.

a case for outstanding results). The rationale behind this particular procedure – which hands the chair of the committee a very important role but is quite usual, as I observed in studies of committees I have conducted elsewhere – was that the presence of the two reviewers prevented me from (voluntarily or not) misunderstanding what they wrote, as they could intervene if I forgot or I over- or under-stated something. More globally, all reviews were also available to all: anyone – with the exception of those with a conflict of interest – could interfere if they thought a reviewer forgot or over- or under-stated relevant information. The publicity of the reviews and of the synthesis I produced guaranteed some legitimacy to the final decision and made it, by definition, collective. Therefore, if neither of the two reviewers protested against my proposal or wanted to add something, and if none of the other members of the committee intervened, the assessment of Mrs Clare's scientific production (to continue with the example) was declared as 'excellent'. Otherwise, a discussion began, and either we slowly came to a general agreement which I again suggested myself after having heard the different positions, or they would vote (I did not) on 'outstanding', 'excellent', 'correct ', 'not sufficient' and the assessment with the larger majority prevailed. In practice, it was rare that there was a vote, which in this case I believe was preferable because the procedure of secret ballot voting that we would use might risk allowing voting for personal revenge or unfriendly votes that cannot be publicly expressed when there is only oral intervention.

Chairing the committee, therefore, plays quite an important role in the evaluation procedure, and is a quite challenging task because of interpretation of the reviews and of the related grades I had to propose during sessions. Building on what was said on the teaching records, the teaching responsibility, the creation of new programmes or the introduction of innovative pedagogy – I had to commensurate all these aspects and produce a synthetic assessment. On top of that, deciding on so many dossiers within two days meant that time was very constrained and long discussions had to be avoided. In order to be as efficient as possible, I spent the whole weekend before the meeting carefully reading all the reviews, looking at the activity reports, and checking the information taken into account by reviewers in their argumentation. Despite the guidelines we sent them, many for instance forgot that only publications published during the period considered for the evaluation were to be taken into account, and sometimes overstated research achievements of those assessed; or they considered papers published in non-peer-reviewed journals as research production while they should be included in the impact domain.

Reading all the dossiers together also revealed some relevant biases. I could detect some reviewers' attitude, especially those finding everything 'outstanding' or, on the contrary those who were never

satisfied. I looked more carefully at the dossiers on which two reviewers had very different views. Of course, I could not pay such attention to everything, but this preparatory work was important to identify those dossiers that might lead to more discussion or controversy. It also gave me an overview of the panel as a whole.

It was always striking to note that reviewers of some disciplines were very generous and others much more critical. The unanimity among economists was especially visible: they all praised the same kind of work and were very laudatory in their comments and arguments. This came from the fact that they had been recruited on very similar basis and all belong to mainstream economics, and that the external members we invited to the committee belonged to the same community – which we were obliged to do if we wanted our economists to be respectful of their reviews – so that they all praised the same kind of research and highly rated all other activities that they don't think are so very important. Political science was at the other extreme of the continuum, as all members of the department do have the same view on what is valuable research and do not agree on a single publication strategy, some being very attached to books while others prioritize papers in peer-reviewed journals. The judgement I prepared for each dossier therefore quite heavily relied on what Karpik (2010) labelled as a personal judgement. I read the arguments with my own knowledge of the tensions and preferences of each discipline, but also with a personal knowledge of the reviewers when I knew them, which was the case for most of them.

When the meeting started, I was able to make a succinct summary of the arguments and a proposed judgement for all domains for all the dossiers. We always worked discipline by discipline in order to try to be discipline-coherent. From this point of view the first dossiers to be evaluated were very important because they set the tone: the level of expectation for the first discipline would then be transferred to the next. This did not mean that the criteria should be exactly the same (impossible to judge dossiers of historians with the same criteria as the economists' dossiers) but that it should be as difficult to be qualified as outstanding in history as in law. So, the first dossiers set the tone but also stabilized the way the committee worked. I became very aware of that after a session where we started with the evaluation of very active professors but who published only papers (in a discipline where books are more than welcome) and reviewers who themselves disagreed on that point. This led to a rather long discussion about the evolution of the discipline and whether the committee should be aware of this trend or fight against it and consider the absence of books as a weakness in the dossier. After that, we were careful to avoid complicated dossiers at the beginning of sessions.

Even if the elements to take into account in the assessment had been defined more precisely, their relative weight in the

commensuration process was still open and often not convergent among committee members. For some, the originality of ongoing research and the complexity of its operationalization should be taken into account and seen to be as valuable as a book while others put more emphasis on publications. The same tension could occur between strong involvement in the institution and strong involvement in the management of international networks, both expected in the third domain but differently valued by committee members.

Trying to avoid such tensions is impossible, unless each and every activity and attribute is codified with a coefficient. No committee would then be needed! As chair, I have to accept that each committee might come to views that are not exactly the same, but at the same time, to be careful of preventing too different criteria across committees, considering that one of the aims was to send rather clear messages to those whose work was evaluated. The memory of what happened during previous committees is therefore important; during the committee itself, one crucial point was to try to be coherent during the two days and to treat the first dossier like the twentieth or the final one. This is the most difficult thing to do. Some of the staff members under my direction assisted me with this task and warned me of potential drift. I also encouraged members of the committee to be aware and tell me if they saw something like that happen. But it did, nevertheless, happen sometimes, quite inevitably.

Quantification of qualitative assessments

This evaluation process is a curious one. It relies mostly on qualitative information, but this information has to be converted into a qualification that becomes a grade, in order to produce a ranking. Three remarks follow about this quantification process.

First, grades lead to overstatement. As described earlier, discussion aimed to reach consensus on the grade to be given to each of the dimensions in the evaluation. But during the last round of evaluations I chaired (the second with the new status and the third since I was vice-president), we decided not to use grades but qualifications ('outstanding', 'excellent' 'satisfactory' and 'unsatisfactory'). We had noted that relying on the A, B, C, D scale led to many As, because, implicitly reviewers considered a B as not valuable enough, or too depreciative. But what we wanted was that an A should be given only for exceptional achievement, and the use of the adjective-scale improved on this. The difference between 'outstanding' and 'excellent', of course, is not straightforward and because what is expected from the work of a colleague may vary from one reviewer to another, qualifications did not prevent divergent views on the same scientific production or on the same teaching involvement. But it nevertheless helped to reach agreement and made for a less skewed

distribution, as opting for 'excellent' was no longer considered as disregarding the work achieved.

Second, it is sometimes difficult to resist the temptation of counting especially about the scientific production. Although the evaluated academics were asked to provide links to their publications, I doubt the reviewers took the time to read them and the written arguments were often such as: 'one edited book, three papers in peer-reviewed journals and four chapters'. The contribution these publications added to the field, their originality or innovative character was rarely mentioned or even argued. I tried to ask for more qualitative assessments or at least ask what was outstanding and not only excellent in the research activity.

Third, the beauty of the thing is that the final qualification was immediately transformed into a number in the computer by one of the staff members under my direction, as 'outstanding' gave 3 points, 'excellent' 2 points and 'satisfactory' 1 point and this excel sheet also directly calculated the final grade once all four domains (three for the researcher-only) had been qualified, with research counting for 50 per cent (80 per cent for researcher only), teaching for 30 percent, institutional and discipline-based engagement for 15 per cent and impact for 5 per cent. This final ranking was therefore progressively computed. It was kept secret, i.e. it was not shared with members of the committee. They could of course use their own excel sheet if they wanted, but I did not see anyone doing something like that. Keeping the final ranking secret was first of all a way of not stigmatizing those at the end of the list or on the contrary valorizing those at the top.

Formative evaluation and merit-based valuation

Development and implementation of this valuation process had two objectives. One of them was what I would call the 'formative evaluation of researchers and professors'. This regular assessment of their activities within the last two or three last years is a way of providing them with an idea of how their work is appreciated, what should be improved, what is expected from them and not there yet. This is the reason why we not only provided grades but I carefully wrote a general assessment of their activity in each domain when we sent them their evaluation. I also expressed recommendations that members of the committee might have formulated. From this point of view, the valuation process is also a policing instrument as it sets the norm of what is considered good or not so good and what should be done. For instance it led some groups to clearly state that this or that journal cannot be considered as peer-reviewed, or to maintain that publishing only in French is not enough, etc. This of course did not immediately lead to a transformation of publishing practices but nevertheless influenced the behaviour of some of those who had been assessed. The publicity of the process (a committee of around 20

internal and external members looking at all the files) probably also affects those who do not want to get 'bad' reviews in front of their colleagues. I also accentuated this formative aspect of the evaluation by inviting those who were assessed poorly for an interview: not to admonish them but to understand the eventual difficulties they met and see how to help them to overcome them. And it worked rather well in some cases. But of course, some resisted and refused to invest in collective services, or still publish in non-peer reviewed journals as the individual cost of the process is rather reasonable.[5]

This evaluation process has, however, another objective, as it is related to performance and remuneration. This means that the result– i.e. the assessment of the quality of X – leads to a valuation process, in the sense that this quality result is converted into a salary increase, something quite different from what I observed for hiring decisions in the US universities in which I conducted interviews;[6] with the process in place in my institution, the highest ranked academics should get a higher salary increase. But as I mentioned before, the budget allocated to the merit-based increase was fixed and limited to fixed percentages of the overall payroll of the evaluated academics. Therefore, from my point of view this valuation phase (how much to increase considering the ranking of each) was the more political moment. Because of the confidentiality of this issue, the decision how to distribute an increase on merit was confined to a small group consisting of a member of my team, one member of the human resources department and myself, and we had to decide how to 'evaluate' the results.

My predecessor's policy was to concentrate this allocation on the very best and to apply the possibility of reaching a 10 per cent increase in salary for only those at the top of the pile, which automatically negatively impacted the potential increase for others as the budget to be distributed was not extensive. As those who are 'outstanding' generally remain 'outstanding' from one evaluation to another, they successively received high increase rates compared with all the others during the evaluation processes that occurred before my own term. Merton's (1968) Matthew effect worked very well. But it meant that those with excellent but not outstanding evaluation were treated like

---

[5] Actually, I did not feel that my colleagues were very anxious about this evaluation, especially if I compare them with colleagues who are on tenure-track, for whom the mid-term evaluation and the tenure process produced much more anxiety as they faced face an in or out decision.

[6] For the recruitment of assistant professors, I observed that quality as assessed by the hiring committee did not play a direct role in the way the price was set afterwards by negotiation between the department chair and the dean. The price of the market (i.e. what is offered by universities considered as equivalent to the recruiting one) was more important than the intrinsic value of the candidate (Musselin 2009 [2005]).

those with lower achievement and received no increase. Thus 'excellent' and 'unsatisfactory' were valuated the same.

In order to avoid the Matthew effect and this absence of discrimination between the less and the better performing, each time our small group discussed what is a 'fair' distribution and where to put the cursor between on the one hand the same increase for all and on the other hand the concentration of a maximum rate on some happy few. What trade-off should there be between expected quality and extreme meritocracy? We opted for a repartition, applying different rates according to the ranking but allowing some kind of increase to more than half the population. We thought it was difficult for someone who received a B (i.e. excellent) on average not to be rewarded and that the highest grades among the 'satisfactory' should also not be completely left out. In other words, the idea was to allocate some rewards to all those who 'seriously' contributed and to smooth the curve. In order to respect the budget, it meant than the highest rate should not exceed 7 per cent. We constructed a first scale with a maximum of 7 per cent for the top of the ranking, 6 per cent for those coming next, 5 per cent etc. But we also took seniority into account: in order to encourage those on the first stage of their career, we retrieved 1 per cent of the given rate for those at the second stage. This means that allocation of merit-based increases did not strictly respect the ranking obtained after the evaluation.

***

Assessing academics, providing evaluations and transforming them into grades leading to a ranking and determining a merit-based salary is therefore a technical activity; but first and foremost it is judgement in action and a political process.

References

Beckert, Jens, and Christine Musselin. (eds.) 2013. *Constructing Quality*. Oxford: Oxford University Press.

Espeland, Wendy N., and Michael Sauder. 2007. "Rankings and Reactivity: How Public Measures Recreate Social Worlds." *American Journal of Sociology* 113(1): 1–40.

Espeland, Wendy N., and Michael Sauder. 2016. *Engines of Anxiety: Academic Rankings, Reputation, and Accountability*. New York: Russell Sage Foundation.

Karpik, Lucien. 2010. *The Economics of Singularities*. Princeton, NJ: Princeton University Press.

Lamont, Michèle. 2009. *How Professors Think: Inside the Curious World of Academic Judgment*. Cambridge, MA: Harvard University Press.

Merton, Robert K. 1968. "The Matthew Effect", *Science* 159(3810): 56–63.

Musselin, Christine. 2009 [2005]. *The Markets for Academics*, New York: Routledge [Le *marché des universitaires. France, Allemagne, Etats-Unis*. Paris: Les Presses de Sciences Po].

Musselin, Christine, and Catherine Paradeise. 2005 [2002]. "Quality: A Debate", *Sociologie du Travail in English*, 47((Supp, 1): e89–e123 [Le concept de qualité: où en sommes-nous? *Sociologie du travail* 44(2): 255–260).

Musselin, Christine, Frédérique Pigeyre, and Mareva Sabatier. 2015. "Devenir professeur des universités. Une comparaison sur trois disciplines (1976–2007) (Becoming a professor. A comparison of three disciplines (1976-2007), *Revue Economique* 66(1): 37–64.

Ossandón, José, Wendy N. Espeland, and Micheal Sauder. 2021 "Islands of Qualities in an Ocean of Quantification": an interview with Wendy Espeland and Michael Sauder, *Valuation Studies,*8(2):103-121.

Christine Musselin is a French sociologist of organizations. She leads comparative studies on university governance, higher education policies, and academic labour markets. She has been a DAAD fellow in 1984-85 and a Fulbright and Harvard fellow in 1998-99. She has led the CSO from 2007 to 2013 and has been the Vice-president for Research of Sciences Po from June 2013 to November 2018. She is the author of *The long March of French Universities*, New York, Routledge (2005) *and The markets for academics*, New York, Routledge (2009). Her last book *La grande course des universités* was published by the Presses de Sciences Po in 2017.