

The French HER system and the issue of evaluation

Catherine Paradeise

UNIVERSITÉ PARIS EST-MARNE-LA-VALLÉE

1. A historical overview

The French higher education and research system is based on a double divide of missions, statuses and recruitment procedures. On one side, over 30 public research organizations – whose permanent scientific staff varies from about 300 to over 12.000 – differentiate themselves from higher education (HE) institutions. On the other, the HE sector itself is made of about 80 universities and about 200 *grandes écoles*.¹ The structure of the system results from the French history of universities, which were suppressed as territorial entities then recreated as a collection of disciplinary ‘faculties’ co-managed from Paris and settled in about 13 big cities during the Napoleonic times. *Grandes écoles* (named *schools* below) were created at the turn of the nineteenth century to educate state engineers on a very selective basis. A school of public administration was added to the list in 1945. Business schools were added in the seventies. Research organizations were built up after the Second World War to face the weaknesses of both types higher education institutions in research. Since the 1990s, various reforms have first incrementally contributed to integrate education and research in both schools and universities, then radically pushed towards building consortia or even entering mergers between higher education institutions.

Although interrelations between these poles developed over time and job contents to a certain extent became more similar, this historical divide remains. Radical reforms occurred at the turn of the twentieth century, which drew these institutions and their evaluation systems nearer to each other but, although they experienced some convergence, they still remain institutionally distinct.

Regarding research units located in higher education institutions, no formal assessment took place outside the “associated research centers” (URA) between universities and CNRS, which emerged in 1965 and largely muted to stronger partnership in “joint research centers” (UMR) between *Centre national de la recherche scientifique* (CNRS) and HE institutions (UMR account for about 90% of all CNRS research centers at the turn of the 2000s). Such units were subjected to a four-year assessment according to CNRS procedures. No other assessment was required, except by ad hoc committees for individual

¹ This number refers to the *grandes écoles* that are accredited by the *Conférence des grandes écoles* among over 400 which deliver post-baccalaureate education.

scholars applying for recruitment and promotion, or public research call for tenders. This state of affairs changed with the radical reforms of the 2000s.²

For its part, not much has changed in the evaluation of teaching. Although rules changed several times since the sixties, recruitment and promotion in universities are basically handled at two levels. First, a national body, the *Conseil National des Universités* (CNU) subdivided into partly-elected, partly appointed disciplinary committees, is in charge of awarding “*qualifications*” to candidates on top of their doctorate (for associate professors) or habilitation (for full professors). Second, local disciplinary ad hoc committees are in charge of recruiting single faculty members (Paradeise ESF 2010). Most *schools* employ a small permanent faculty and a large number of high-level adjuncts who are hired on short-term contracts while being permanent members of universities, research centers, administration, business and industry. As in universities, the academic staff is usually not assessed after recruitment, except for promotion. Evaluation occurs at recruitment and promotion only. Should they not fit the needs of the school and students’ evaluations, their contracts would not be renewed. Research organizations are the only institutions, which carry out a periodical formal assessment since their foundation. They evaluate both their research centers and their full-time researchers, usually on a four-year basis, based on partly peer-elected, partly appointed committees in each large disciplinary field.

As elsewhere in Europe and in many countries worldwide, radical reforms followed, starting 2006, the incremental phase of the 1980-90s. The issue of evaluation had already been considered in the 1970s and more and more in the 1980s. Yet the need to assess all academics, universities and research centers only became consensual in the 1990s (Merindol 2008). Consensus developed as a counterpart of a rising awareness that more autonomy would benefit all stakeholders of universities, which were still highly dependent from state authorities. In 1983, four-year contracts on research and teaching were first introduced between the state and each university, bringing the latter to be identified as an assessable organizational body of its own, able to strategize and plan its future and to argue for its funding application. The *Comité National d’Évaluation* (CNE) was set up in 1985 – with little resources and major ambitions – to improve transparency on HE institutions performance. An *Observatory of Sciences and Techniques* (OST) was founded in 1990 to forge indicators of performance, with the purpose to progressively better support allocation decisions.

The 2006,³ 2007,⁴ and 2013⁵ legislative acts ruled on the autonomy and accountability of universities. Although it was at first perceived as a major break-through, it only granted a limited autonomy as compared to other European countries (EUA 2011), but it did impel the foundation of systematic tools and methods backing assessment on performance, and allocation on assessment.

² CNRS is the largest basic research organization. It is followed, in term of scientific staff numbers, by two targeted research centers, the *Institut national de la recherche agronomique* and the *Institut national de la recherche médicale* (INSERM).

³ *Loi de programme* No. 2006-450 du 18 avril 2006 pour la recherche.

⁴ *Loi* No. 2007-1199 du 10 août 2007 relative aux libertés et responsabilités des universités (so-called LRU).

⁵ *Loi* n° 2013-660 du 22 juillet 2013 relative à l’enseignement supérieur et à la recherche (so-called loi ESR).

On the one side, a national funding agency, *Agence nationale de la recherche* (ANR) was set up in 2006 with the purpose to increase the share of competitive public funding of research, partially based on research programs and partly on open programs. Moreover, the creation of a *Commissariat aux investissements d'avenir* (CIA, General Commission for future investments) in 2012, stressed the importance of large competitive funding of many institutional and operational levels of excellence consortia of universities (IdEx), laboratories of excellence (LabEx), excellence facilities (EquipEx), etc. This investment program, which starts its third round in 2017, has already dedicated 22 billion euros to “excellence initiatives” in higher education and research across the country.

On the other side, public authorities set up in 2007 an evaluation agency, *Agence d'évaluation de l'enseignement supérieur et de la recherche* (AERES) renamed in 2013 *Haut comité à l'enseignement supérieur et de la recherche* (HCERES). Its main duty is to assess all components of public higher education in all fields and at all levels outside individual scholars (HE institutions, research organizations, research centers and degrees). A (poorly) called “SYMPA formula” was elaborated at the same time by the ministry bureaus to translate performance in terms of funding allocation. It was intended to base a proportion of universities block grants (about 20%) on their outputs in teaching and research, in addition to the 80% based on their inputs (number of students in various fields, etc.). The internal allocation of university block grants was to be handled by the single universities themselves.

2. Reception and impact of assessment

The accountability turn accelerated the production of tools as a basis for resource allocation. In addition to the introduction of cost accounting, autonomous universities created their own management dashboards to back their strategic decision-making. The evaluation agency disseminated its own lists of detailed indicators on universities, research centers and curricula,⁶ in order to assess their organization, governance, funding and performance.⁷ Indeed such indicators inform the assessment, which is yet based on the evaluation drawn up every four years by ad hoc visiting committees using basically few metrics. As far as universities are concerned, such committees are required to synthesize their evaluation under a set of dimensions, covering governance, leadership and management, and strategy in research, knowledge transfer, teaching, student life, external and international relationships and communication. Curricula are assessed according to their goals, context, organization and results. Research centers are assessed on six dimensions: scientific production and quality, reach and attractiveness, interaction with the social, economic and cultural environments, internal organization and life, involvement in teaching research, and strategy and five-years programme.

⁶ More units are assessed, where for instance federal structures or consortia of universities (and possibly *schools*) have been set up.

⁷ Such as number and status of academic and management staff, organizational chart, decision-making procedures, etc.; attractiveness and placement of curricula; grants captured, patents and publications (based on a list of ranked refereed journals built up by disciplinary ad hoc committees), etc.

2.1. Reception of assessment and its consequences on assessment tools and processes

IS THE ASSESSMENT AGENCY LEGITIMATE?

A heated debate first developed about the issue of accountability. Each discipline used its own channels of influence to seek better arrangements for itself. Strong and politically threatening lobbies such as *Sauvons la recherche* brought together individual and collective discontent, based on shared protest against the so-called neo-managerial or neo-liberal turn in higher education. Such pressure groups resisted the use of assessment tools as a basis for funding, and they claimed the need to adjust evaluation criteria to the specificities of disciplines. The most radical individuals and groups disseminated the recommendation to boycott invitations to join ad hoc committees or to resist English speaking in such committees. At the end of the day, they organized political pressure to get rid of the agency and its tools.

Collectively, the attractiveness of protest varied depending on how much assessment was individually and collectively considered as a threat. Collectively, the less accessible the fields to international journal rankings – either because they belong to “non-nomothetic fields” such as the social sciences and humanities (Passeron 1991) or because they deal with professional knowledge as in law studies, accounting and engineering prone to pragmatic knowledge rather than academic publications – the more threatened they felt. Individually, the more scholars cumulated disadvantages such as being low-publishers and having lower academic statuses, the more they were likely to reject the assessment-driven model. The more disturbing reforms for the social exchange model they had built up with their own university, the more they were likely to raise their voice (table 1).

Table 1. Contributions of scholars, expected returns and reception of assessment.

Exchange model between scholar and...	Individual commitment	Expectations towards one's university	Individual behavior
...his/her university	Involvement in organizing and managing	Local promotion should reward organizational commitment	Withdrawal if HRM does not deliver promotion. Hostility towards the notion and measure of scientific excellence and voice against assessment and performance-based resource allocation
...his/her discipline	Personal conquest of excellence	The university should provide resources for personal accomplishment	Exit if the university does not comply with expectations, based on demonstration of an excellent performance. Positive behavior towards the notion and measure of scientific excellence and performance-based resource allocation
...both	Both are considered complementary	Local promotion should reward scholars committed as academics and as good citizens of the university	Local withdrawal or exit if the university does not meet the expectations. Voice in favor of pushing forward the reforms in such a way that universities really become complete organizations

The assessment frame evolved over time under such pressures, but the general rationale of the reform remained untouched. Indeed the assessment agency (AERES) was theoretically discontinued to comply with lobbies during the 2012 presidential campaign, but was practically immediately reopened under the 2013 ESR act with a new name (HCERES) and with a slightly renewed legal status, which kept though almost exactly the same jurisdiction as before.

WHAT SHOULD BE ASSESSED AT THE NATIONAL LEVEL?

One first struggle about assessment concerned the selections of entities to be assessed by AERES/HCERES. Each institutional actor – research organizations, the university system and *schools* – argued that it already implemented its own individual assessment. Their coalition, backed by individual scholars, led to finally leave them out of reach. The same fight occurred on research centres, but it was unsuccessful. It was indeed a major stake for the government to arrange uniform tools allowing for the development of research incentives, whether or not associated with any public research organizations. By linking part of block grants to the university outputs in research as measured by formal indicators such as number of patents and rates of publications in so-called best international journals, it was supposed to encourage individual scholars to improve their own performance and good research centers and universities to select best-performing scholars.

The development of research performance-based funding, however limited it actually remained, disrupted the traditional allocation scheme, by making it clear which centers and universities concentrated good publishers and which did not, thus reconsidering their bases of reputation (Paradeise and Thoenig 2015). Many scholars – individual academics or subgroups – felt they might be stigmatized as low research-performers. They also feared that such uniform indicators, which might not fit their ways of publicizing their research, would erode their disciplinary and cultural specificities and ostracize their field, by for instance favoring journals against books or memoirs, and English-written publications against their native language. Such reactions mostly took place among low- or non-publishers, and among humanities and social scientists.

SHOULD ASSESSMENT BY MADE PUBLIC AND HOW?

Making assessment reports and ratings public was a major change introduced with the creation of AERES. Used to shame or, rather, legitimize policies and funding, no one in the same field or in the same institution could ignore the comparative performance of research units or institutions. This would allegedly help the state decide upon resource allocation across universities, universities decide upon resource allocation between its sub-units, research groups and departments strategize⁸ in order to try get rid of non-publishers or improve their scores through better recruitments, etc. Academics feared a mechanical implementation of scoring on decision-making, while management dashboards could be used in many other ways, for instance to reinforce a poorly performing discipline which, for some reason, was considered important by a given university. Protest denounced the illegitimacy of such publicity. For these reasons, they first opposed synthetic scores on a scale of five, (from A to E) which the visiting committees were required to deliver. Soon after the first round of assessment, the agency decided to buckle under this pressure and reconsidered ratings as a list of itemized non-additive scores on each of the dimensions under evaluation. Finally, it was invited to totally renounce and frankly discontinue any form of scoring.

⁸ We have shown elsewhere how much the strategizing capacity varies from one place to the other within a single country (Thoenig and Paradeise 2016).

HOW SHOULD JOURNALS BE ASSESSED?

One important issue emerged that was about the ranking of journals in certain fields, which feared inadequacy of the uniform criteria applied in assessing publication performance (table 2). “There is a classification of journals in natural and life sciences... (which is) mostly based on English-written journals. Such a classification does not exist in the social sciences and humanities. And it seems to poorly fit academic outputs in these fields, largely French-written” (Glaudes 2014). As a result, the behavior of disciplinary communities varied widely about the injunction to list reference journals in their field and even rate them on a three steps scale.

Table 2. Disciplinary publication cultures.

	Math	History	Political Science, Sociology	Chemistry, Biology
Privileged medium	Journals, Annals	Books, Conferences	Journals, Books	Journals
Agreement on media hierarchy in the community	Strong	Average +	Average -	Strong
Production time	Slow	Slow	Medium	Rather fast
Citation accumulation	Very slow	Medium	Medium	Fast
Citations obsolescence	Slow	Slow	Slow	Rather fast

Certain disciplines, such as economy or management, simply replicated lists built elsewhere, for instance by public research organizations or international newspapers such as the *Financial Times*. Others, such as law, built up their own list, based on the empirical signals of reputation established by their representative authorities. Others used their own indicators without any effort to link them with those of others. Philosophers defined for instance their own four specific criteria (requirement to principally publish articles in philosophy, existence of a scientific committee and an editorial board, including non-French members, double-blind evaluation, selectiveness). On top of similar criteria, communication scientists added up a list of other items such as regularity of publication and size of articles, restriction of auto-publication, institutional links with the discipline and indexation in international databases.⁹

Methodological diversity added up to differences in the established lists. Several disciplinary committees in the field of arts, social sciences and humanities (for instance in sociology, political science, theology, philosophy, anthropology, geography and urban planning, history, arts and law studies) simply refused to set up journal rankings. Some contributed by listing journals that belonged to their scientific perimeter. They promised to, and did progressively develop their own rankings (communication studies, psychology), each with its own scale. Finally, a series of disciplines (concentrating in languages, literature and civilizations) totally rejected the very notion of a list, arguing that they were irrelevant in their field.

The AERES finally took notice of this resistance and, since 2010, started rebuilding the lists, including other items based on a more cautious typology of media, such as scien-

⁹ See online <http://www.aeres-evaluation.fr/Publications/Methodologie-de-l-evaluation/Listes-de-revues-SHS-de-l-AERES>.

tific books (based on publishers, signatures, purposes and editorial work). It made it clear that it did not favor quantitative evaluation but followed the moderate recommendations of the French academy of science¹⁰ by more generally restricting the use of bibliometrics to the assessment of entities which size exceeds 30 scholars, and systematically referring bibliometric results to their average values and to the 10% top values in a given field. In addition, experts – who are always peers in the fields under assessment – were invited to be cautious about possible biases of such results. Thus AERES fostered a non-mechanic use of bibliometric indicators and insisted that they should be contextualized and interpreted, and should not replace the reading of papers in order to assess their actual scientific interest.

WHAT SHOULD BE THE PERFORMANCE OF A “PUBLISHER”?

Lobbies paid attention to the norm set up to define what to be a “publisher” means. They also insisted that this norm should vary according to the publication tradition of each field. They worked at lowering the threshold and finally ended up in the social sciences and humanities accepting a (very) light norm of 4 articles for a full-time researcher or 2 articles for a professor in a four-year period, with indeed very little variation from one field to the other.

OVERALL STRENGTH OF ASSESSMENT IN FRANCE

To tell the truth, protesters over-emphasized the threats of assessment, at least during the current stage of reforms. On the one hand, a recent EUA survey shows that the impact of evaluation of teaching is comparatively very low in France (table 3). On the other hand, research outputs are regularly assessed, have gained influence on recruitment and promotion, but have no impact whatsoever on tenure and salaries of academics who remain civil servants paid according to a fixed national grid of statuses.

Table 3. How much is assessment actually developed in France? (source: EAU)

	Europe	France
Students satisfaction survey are used in the assessment of teachers	93%	39%
Courses are regularly assessed	89%	33%
Research outputs is regularly assessed	84%	100%
There are ways to face situations where teaching performance of someone remains weak	72%	18%
Heads of departments/deans of faculties discuss regularly the quality of teaching with all teachers	67%	24%

In addition, the impact of evaluation by the assessment agency on resource allocation remains limited. Yet, other forms of evaluation play a major role in differentiating individuals, research centers and universities in the competition for grants.

¹⁰ See online <http://www.academiesciences.fr/activite/rapport/avis170111.pdf>.

2.2. The impact of assessment on funding

THE IMPACT OF AERES/HCERES

Performance evaluation by assessment agencies was developed as a tool of accountability for a better governance of the new autonomous universities. It justified building all sorts of indicators intended to inform the SYMPA formulae mentioned above. This top-down tool however remained ineffective until 2017. First, it took time to set up the required databases. Second, many universities considered that a top-down approach to the building of indicators would not be able to capture their actual performance. Third, in a context of stability and even reduction of higher education public budgets, the government did not dare bypassing the established distribution of block grants. In other words, the top-down approach, which prevailed in the SYMPA formulae as a link between performance and allocation, was discarded almost as soon as it was created. Starting in 2016, a new bottom-up approach by sector, developed with the collaboration of HER institutions, is supposed to design a new formula that should better fit the specificities of each field.

To put it in a nutshell, the output-based assessment by AERES/HCERES has since its foundation proved rather ineffective operationally. Nevertheless, its symbolic impact has been enormous by making publicly known the strengths and weaknesses of units and sub-units in the system of HE and research, by fostering strategic moves at each level, by setting up the issue of publication and by insisting on its contribution to the missions of academics, etc.

THE IMPACT OF ANR AND CIA

On the contrary, the development of project-based grants over the last ten years has had major operational impacts on the dynamics of universities, research centers and to a certain extent individual careers, first with the funding of research projects by ANR programs and increasingly with the substantial sums supplied by the CIA institutional excellence initiatives. Three waves of funding have been set up since 2012, covering a variety of large projects involving not only research but also the founding of new institutional bodies – laboratories of excellence (LabEx), excellence facilities (EquipEx), excellence institutions (IdEx). The CIA program progressively diversified its funding streams, which now include innovation in teaching as well as incentives targeting the development of specific niches of excellence within universities with the *Initiatives Science-Innovation-Territoires-Economie* (I-SITE). The international high-level evaluation committee pays much attention to the relevance of projects in scientific and operational terms but also to their feasibility in terms of governance. The important resources procured by such programs operate as very strong incentives that also encourage the development of project-based consortia and even mergers between research centers, departments, universities and schools. The CIA programs have thus come to play a key role in the current on-going stratification of French higher education and the restructuring of the national landscape, both at the institutional level and between and within disciplines. By concentrating important resources on specific territories, these programs favor the visibility and attractiveness of certain universities or certain niches within universities.

2.3. The impact of assessment on the profession

The development of assessment has provided a rationale for the redistribution of resources between universities and between disciplines. It may not have had much impact upon individual salaries and national careers of academics, who remain civil servants, but it has positively changed working conditions of the best-performing units, whatever the discipline. Altogether, humanities and social sciences have received less budgetary resources than hard sciences, partly because they display lower needs than experimental sciences, partly because assessment tools too often have difficulties grasping their specificities. But, as shown by the relative growth of their memberships, being in line with the massification of higher education in a non-selective system, they have not been ostracized as such (table 4).

Table 4. Size of permanent academic staff by discipline. 1992-2013 (source: DGRH, French Ministry of Higher Education and Research), mentioning the position of political science and the highest and smallest growth in each large field.

Field	2013	% Growth 1992-2013
Social sciences and humanities	14,999	+57.2%
<i>Sociology</i>	877	+98.9%
<i>Literature and languages</i>	6,132	+24.0%
<i>Political science</i>	360	+59.3%
Hard sciences	24,907	+40.7%
<i>Engineering</i>	6,845	+71.4%
<i>Math and informatics</i>	6,601	+63.9%
<i>Physics</i>	2,429	+2.0%
Medicine and pharmacy	8,050	+2.4%
Grand Total	55,515	+40.4%
Students population	2,471,000	+43.9%

Yet, political scientists have been impacted, as have other scholars especially in the social sciences and humanities. The development of performance-based assessment, however limited if compared to some other European countries, has revealed a more visible hierarchy between scholars. Reputations and statuses have been tested by performance as measured by “excellence” metrics (Paradeise and Thoenig 2015). The worldwide generalization of accountability is segmenting the academic population, building up a pecking order between first-class and second-class scholars, publishers and non-publishers, members of top, second- and third-tier institutions. As stratification between universities increases, one may expect that the best-rated departments and/or universities will increasingly attract first-class scholars, who are also chased on the international market and whose salaries may become much more flexible and substantial. Two labor markets are thus being created. Roughly speaking and with several exceptions, the international one increasingly takes care of the “stars” while the national one takes care of the others. Since French civil servants’ salaries are all but competitive, institutional reputation will not be enough in the future to prevent more academics to leave the country, a tendency already confirmed by a still limited but increasing trend among younger scholars.

3. Conclusion: pros and cons of research assessment

When considered at the systemic level, the obsession of HE policies to make French universities “visible from Shanghai” could endanger universities and departments which have no hope of accessing the Walhalla of world excellence, but place a major emphasis on the higher education of large segments of the young population. Thus, France should remain cautious not to concentrate evaluation solely on cutting-edge research. As other European countries, it should take care to preserve and encourage the many and varied “excellences” that are needed to face the various missions of universities.

For all these reasons, it is difficult to sum up and provide a uniform overall assessment of research assessment in France. The reception of assessment mostly co-varies with the opportunities it provides and the threats it involves for universities, faculties, research centers and individuals. The analysis of such opportunities and threats does not identify disciplines and scholars that would uniformly be the losers or the winners of the new rules of the academic game but rather cuts across all of them (table 5).

Table 5. Reception and impact of assessment in France. A synthesis.

	University level		Professional level		Individual level	
	Pro	Con	Pro	Con	Pro	Con
Reception	Backs strategic perspectives	Fear that strategic decisions be taken in a neo-managerial perspective and hurt collegial regulations and the very nature of universities	Resource for challengers as opposed to oligarchic cliques within the profession	Appropriateness of metrics varies across disciplines Finalized competitive grants may negatively impact blue sky research	Fragmentation between publishers and non-publishers	Fragmentation between publishers and non-publishers Publications and patents do not account for all types of contribution as a scholar
Impact	Contributes the rising stratification of universities by differentiating resources: recruitment, budgets and facilities. Strategizing involves more conversation between departments to decide on and legitimize the allocation of resources.	Contributes the rising stratification of universities. Risk to try get rid of disciplines that do not score in rankings and bring costs rather than budgetary resources. Possible emergence of internal political fights based on divergences between departments on what must be valued in the academic work	Criteria of recruitment and promotion necessarily involve research activities Contributes the rising stratification between departments in a given field by differentiating resources: recruitment, budgets and facilities	Contributes the rising stratification between departments Risk of constraining recruitment and promotion choices based on performance as defined by metrics	Rewards who does the academic job (teaching, research, administration). Encouragement to target better vectors and upgrade publications	Disqualification of non-publishers, even when they do a good job in teaching

References

- Du bon usage de la bibliométrie pour l'évaluation individuelle des chercheurs, Académie des sciences, janvier 2011.
- Glaudes, P. 2014. L'évaluation de la production scientifique en France par l'Agence d'évaluation de la recherche et de l'enseignement supérieur?, *Mélanges de la Casa de Velázquez*, 2, Tome 44, 293-300.
- Merindol 2008 Mérimond J.-Y. 2008. Comment l'évaluation est arrivée dans les universités françaises, *Revue d'histoire moderne et contemporaine*, 55-4bis, 7-27.
- Passeron J.-Cl. 1991. *Le raisonnement sociologique. L'espace non-poppérien du raisonnement naturel*, Paris, Nathan Essais et Recherches.
- Paradeise C. 2011. 2011. Higher education careers in the French public sector, Permanence and change, in Avedduto S. (ed.), *Convergence or differentiation. Human resources for research in a changing European scenario*, ScriptWeb: Napoli, 159-184.
- Paradeise and Thoenig 2015. *In search of academic quality*, London: Palgrave-Macmillan.
- Thoenig J.-C. and Paradeise C. 2016. The strategic capacity of academic institutions, *Minerva*, 54(3), 293-324.