TABLE OF CONTENTS
MEHEM (Mapping European Higher Education Models)
EU Sponsored Project number TR 0604.01-03/44

Comparative Study of Performance Evaluation Systems in European Higher Education:
Mediterranean Models

Draft version

Severine Louvel

1. Introduction: theoretical framework – four types of research evaluation systems........... 1
2. Spain................................................................................................................................. 4
   2.1. Since 1989: an intrusive and weak research evaluation system to evaluate individual
        researchers................................................................................................................ 4
   2.2. Since 1987: limited developments of the evaluation of research organizations ...... 5
3. Italy............................................................................................................................... 7
   3.1. Since 2001 the “Italian Research Assessment Exercise” or the systematic evaluation of
        public universities and research institutions ............................................................. 7
4. Turkey: the limited institutionalization of the research evaluation system...................... 12
   4.1. Research evaluation activities at the organisational level....................................... 12
   4.2. Faculty performance evaluation ............................................................................ 13
5. France.......................................................................................................................... 15
   5.1. The evaluation of public research institutions: from “weak and intrusive” to “strong
        and intrusive” and/or “strong and competitive”? .................................................... 15
   5.2. The evaluation of individual academics: still “weak and intrusive, but more
        consequences on career progressions....................................................................... 16
6. Synthesis....................................................................................................................... 17
7. References ...................................................................................................................... 19

1. Introduction: theoretical framework – four types of research evaluation systems

Evaluation, quality and performance assessment in higher education have become important issues in all European countries. Within a very large set of questions and problems (evaluation of teaching, ranking of universities, evaluation of faculty members, of research organisations, etc.), this report focuses on the evaluation of research activities, be it at the individual or and the organisational levels. Its aim is to collect data on the research performance evaluation procedures which have developed over the last decades and which are currently employed in the four Mediterranean countries studied in the Mehem project: France, Italy, Spain and Turkey. The report will develop ideal-typical descriptions of research performance evaluation regimes in these four countries. While emphasizing similarities and differences between the countries, the report will serve as an input into a broader

Severine.louvel@iep-grenoble.fr
IEP - PACTE (CNRS / Grenoble Universities) - 1030 avenue centrale, Domaine Universitaire, 38400 Saint Martin d’Hères, France.
assessment that will attempt to capture diversity and homogeneity within European performance evaluation systems in higher education.

The report is based on existing literature on the topic, on the Mehem country reports and on presentations made at the Mehem workshops, and finally on additional online information (e.g. institutional websites of ministries, evaluation agencies, etc.). However, due to the lack of empirical information and/or to the very limited development of performance evaluation procedures in some countries (e.g. Turkey), this report is not always able to provide an in-depth analysis.

In order to compare the four countries studied, the report will draw on the four archetypal models of performance evaluation developed by J. Gläser (Gläser, 2007). The author differentiates the research evaluation systems (RES) along two dimensions: the type of information they gather and provide (intrusive versus competitive systems); the strengths of their consequences for the individual researchers or for the research organizations (weak versus strong systems).

<table>
<thead>
<tr>
<th>Information provided</th>
<th>Consequences weak</th>
<th>Consequences strong</th>
</tr>
</thead>
<tbody>
<tr>
<td>Need for improvement (local, intransparent information)</td>
<td>Intrusive research evaluation systems</td>
<td>Hierarchical intervention</td>
</tr>
<tr>
<td>Comparative position (public, transparent information)</td>
<td>Competitive research evaluation systems</td>
<td>Competition for funding, reputation, and students</td>
</tr>
</tbody>
</table>

**Intrusive RES** rely on qualitative in-depth information, mostly gathered by peer review or university self-assessments. As they are informally organized and as the information gathered is mostly used informally, they can be qualified as “non transparent” and “not public” (Whitley, 2007). Such information usually enables a low level of comparability. In deed its aim is not to stimulate competition between individuals or organisations while ranking them according to their respective performance. Evaluation procedures can take place ex-ante or ex-post.

When *weak*, intrusive RES do not have consequences on the funding of research: the evaluation outcomes take the form of recommendations, which are usually highly specific and contextualized, and which may enable the research organisations or the individual researchers to improve the conduct of their research. As they primarily intend to improve research capabilities they can be described as “formative” (Whitley, 2007).

When *strong*, intrusive RES do have an impact on the funding of research (or, in case of the evaluation of individual researchers, on their salaries or career progression).
**Competitive RES** rely on qualitative or quantitative information, which is gathered by peer review or by indicator based procedures. Contrary to intrusive RES the evaluation procedures are here formally organized and standardized and the evaluation outcomes are highly comparable. In this sense competitive RES can be described as “transparent”. Moreover, the information on evaluation is publicly available: in deed the aim of such RES is to foster competition between research organisations of between individual researchers.

*Weak* and competitive RES have no or weak consequences for the allocation of research funding. In these cases the State carries out the evaluations, thus creating competition between them, but it lets universities being responsible for improving their research performance. This may be seen as a result of the institutional autonomy of universities. The various rankings produced by non State actors may be included into “weak and competitive” evaluation procedures.

*Strong* and competitive RES are consequential for funding. The allocation of research funding is based on the research performance and the RES creates incentives to improve it.
2. Spain

In Spain, the institutionalisation of a RES goes back to the mid-1980s (Cruz-Castro and Sanz-Menéndez, 2008). This institutionalisation took place at three levels: 1) an increase in the use of peer-review in funding allocation decisions, leading to the emergence of peer review project-based funding regime and to a strong steering of research by funding bodies; 2) the establishment of a model for the evaluation of individual researchers; 3) limited attempts to establish evaluation procedures for research performance at the organisational level.

2.1. Since 1989: an intrusive and weak research evaluation system to evaluate individual researchers

The case of Spain is quite specific in Europe as research activities are mostly evaluated at the individual level (faculty members and researchers of the CSIC – the largest Spanish research organisation with over 10,000 employees and 2500 tenured scientists). The evaluation procedures of individual academics were introduced in 1989 with the LRU University Reform Law (Ley de Reforma Universitaria). Whereas universities were in charge with the evaluation of teaching activities, a newly created National Commission for the Evaluation of Research Activities (CNEAI) conducted the evaluation of research activities. The CNEAI is composed of experts appointed by the Ministry of education and science and of representatives from the regional governments.

The evaluation procedures, as well as their effects on the Spanish research system, have been extensively analysed by (Jimenez-Contreras and Delgado López-Cózar, 2003). This report will here summarize their main developments and conclusions.

The CNEAI was established in order to increase the university professors’ productivity by giving them economic motivation to do so. Information on the individual performance was mostly gathered by peer review and the procedures followed explicit criteria. In deed, the evaluation gave preference to the publication of work in international journals listed in the ISI’s Journal Citation Reports. Nevertheless, the fields of Social Sciences, Humanities and Judicial Sciences were evaluated according to alternative criteria (books, etc.). Once a year, tenured researchers and faculty staff submit – on a voluntary basis – five contributions which are evaluated by a panel of experts.

The maximum increase in salary that researchers can get after 6 years evaluation period is equivalent to less than 3% of the average yearly salary of titular university teachers. Consequently, the symbolic rewards coming with a good evaluation may be even more important than economic ones. As stated by (Cruz-Castro and Sanz-Menéndez, 2008), the use of standard criteria – and in some fields of bibliometric

---

2 These ex-ante evaluation procedures will not be analysed here.
3 There are 11 scientific fields: Mathematics and Physics, Chemistry, Cellular and Molecular Biology, Biomedical Sciences, Life and Earth Sciences, Engineering and Architecture, Social/Economical/Political/Behavioural/ Educational Sciences, Economics and Business Studies, Law and Jurisprudence, History and Art, and Philosophy/Philology/Linguistics.
indicators – have had a strong indirect impact on the informal “reputational market” of academic science. Moreover, it has induced clear changes in the publication pattern of Spanish academics (with an increase in production rates since the 1990s and a growing share of publications in international journals). Finally, even though the evaluation of individual performance is not consequential for obtaining funding for research, it is widely considered that having received a bonus through the CNEAI evaluation increases the chances of being selected in the competitive bidding processes of the funding agencies.

To summarize the findings about the Spanish research evaluation regime for individual academics, Gläser considers it as a “weak intrusive RES” (Gläser, 2007). In deed, it has weak consequences for the salaries of the assessed academics. Moreover it relies on qualitative information which is mainly used for internal purposes, that is, to increase the individual researchers’ performance. One can however question if there is not a limited shift from an “intrusive RES” to a “competitive” one, due to an indirect use of the data collected on individual research performance. In deed, as stated by (Cruz-Castro and Sanz-Menéndez, 2008), the data were also used in an aggregated form which has begun to affect collective reputations. More precisely, institutions such as the Ministry for higher education have started in 2004 to publish some data about the research performance of universities which are similar to rankings.

Beside the CNEAI evaluation policy, one should also notice that the 2001 Universities act has introduced new evaluation procedures which are also intrusive insofar as they are directly linked to the recruitment and promotion procedures of universities (Cruz-Castro and Sanz-Menéndez, 2008). In deed, access to some pre-tenured teaching and research positions is now conditioned by an accreditation procedure (evaluation of teaching and research activities by a national agency). Access for tenured positions (with civil servant status) is also conditioned by an habilitation procedure (also carried out by the National Agency for Quality Assessment and Accreditation created in 2001 – ANECA).

2.2. Since 1987: limited developments of the evaluation of research organizations

As mentioned previously, the Spanish research evaluation system which has been institutionalized since the mid-1980s at the national level as well as at the regional one is focused on the individual researcher and on the ex-ante peer-review evaluation of research projects by funding agencies. By contrast there has been only a marginal interest in the evaluation of research performance at an organizational level (research groups, universities, etc.). However, there has been limited developments regarding organizational assessment which are analysed by Cruz-Castro and Sanz-Menéndez (Cruz-Castro and Sanz-Menéndez, 2008).

First, a National Agency for Evaluation and Foresight (ANEP) was created in 1987 by the inter-ministerial commission for science and technology (CICVT). It had three main attributions: project evaluation (requested by funding bodies); general assessment and technical advice to political bodies; and evaluation of research
organisations. However evaluation of research organisations remained limited to single cases, so that up to the mid-nineties “organisational evaluation directed to strategic purposes constituted an exception in this context, where it appeared only under very exploratory actions with little practical consequence” (op. cit.)

Second, the National Plan for Quality Assessment of Universities (PNECU) which started in 1996 with three following rounds (1998, 1999 and 2000) was designed to promote the institutional assessment of university quality on the basis of homogeneous methods. The plan clearly indicates the institutionalization of a “weak intrusive RES” as it relies on qualitative information (obtained by a mix between self- and external assessment; synthesized in a report) and as its aim is to help internal and external decision makers. Furthermore it was not consequential for funding. The plan was supposed to evaluate teaching and research activities (scientific production, but also funding, management, etc.), however its main focus remained teaching\(^4\) rather than research.

Finally, many regional governments have created accreditation and quality assurance agencies after the University act in 2001. Although these agencies are also commissioned with the evaluation of research programs and organisations, these activities are rather underdeveloped compared to the evaluation of teaching programs and that of individual researchers. Moreover this evaluation procedures can also be considered as “intrusive and weak” as “in almost no region has a 'research performance evaluation' been linked to the funding regime of universities and public research organizations” (p. 219). Consequently research evaluation is not considered yet as a tool for strategic planning and management improvement.

Cruz-Castro and Sanz-Menéndez (ibid.) give two main explanations for this underdevelopment of research evaluation activities at the organisational level. First, the rapid expansion of undergraduate education during the last two decades may explain why the evaluation of teaching had to a certain extend priority over that of research. Second, the multiplication of funding sources since the mid-1990s was responsible for the dramatic growth of the research system, which in turn contributed to the overload of the national system for evaluation. However, the multiplication of regional evaluation entities partly compensated for this overload.

\(^4\) According to Cruz-Castro and Sanz-Menéndez, the plan was quite successful in this regard as it introduced a culture of evaluation within universities, and as it helped improving their information and statistical systems.
3. Italy

3.1. Since 2001 the “Italian Research Assessment Exercise” or the systematic evaluation of public universities and research institutions

In Italy, the first systematic evaluation exercise of research activities was completed between 2001 and 2003. It is fully described on the website of the committee for the evaluation of research (CIVR) ([http://vtr2006.cineca.it/index_EN.html](http://vtr2006.cineca.it/index_EN.html)) and thoughtfully analysed by Minelli, Rebora and Turri (Minelli et al., 2008).

Evaluation activities in Italian universities have traditionally favoured a bureaucratic approach based on the ex ante assessment of compliance with legal and administrative rules.

The committee for the evaluation of research (CIVR) was set up in 1998. The exercise, known as VTR, regards research activities carried out between 2001 and 2003. The exercise aimed at evaluating the scientific performance of 77 universities (both State and non State), 12 State research institutions and 13 other research institutions (these latter participated willingly). The results were published in 2006. The CIVR exercise represents the first national evaluation of university and non-university research in Italy.

Procedures

Each university or research institution selects research products corresponding to 25% of the total number of permanent academic staff. Consequently the CIVR exercise does not assess the overall quality of the university research. It rather evaluates the ability of universities to produce a certain number of research outputs of international excellence.

The research outputs are transmitted to the CIVR panellists (20 disciplinary panels) with a maximum of 150 research outputs by panellist. However the evaluation is not performed by the panellists themselves but by external experts who are collectively appointed by the panels. Each research output is assessed on a 4-dimensions scale (quality, relevance, originality/innovation, international outreach). The panels then synthesize the evaluations in a single-scale rating (excellent, good, acceptable, limited). Finally, each university is given two ratings for its research performance (the mean of the scores obtained by its products selected in the areas; the percentage of excellent products). The ratings posted on the CIVR website group the results according to the number of selected products (megastructures with over 74 products; large structures between 25 and 74; medium-size structures between ten and 24; and small structures with under ten products).

The CIVR exercise presents the first overall picture of the results of research activities in Italy. Minelli et al. (2008) consider it an institutional innovation in several ways:

- The information provided is public: the CIVR posted the rating obtained in 20 disciplinary areas by all universities on its website and placed the universities’ performance in a ranking list. The university ranking lists become an element of accountability that is also available to the broader society (students, enterprises,
other universities). However detailed information for each university (that is, the rating of individual research products) is only communicated to the rector and in a synthesized form. The fact that individual researchers were not told their own ratings was considered a limitation of transparency. Consequently we should add here, in order to relate the VTR to the theoretical framework on evaluation describe above, that the Italian RES is competitive rather than intrusive, as its explicit purpose is to rank universities according to their research performance in order to introduce some competition between them.

- The ratings are based on ex-post performance rather than on ex-ante compliance with rules and parameters. The VTR changes the perspective of Italian evaluation by “leaving behind the ‘state control’ model of higher education based on centralised decision and bureaucratic control with its attention to formal aspects, inputs and compliance with standard processes. The CIVR exercise, on the other hand, enhances the ‘evaluative state’ model based on autonomous universities and controls of performance” (Minelli et al., p. 233)

- The methodology which relies on the active participation of the scientific community (as it has to chose the research products which will be evaluated) is quite innovative. Nevertheless it should be noticed that it somehow extends the peer review process Italian academics are used to. It that sense it does not introduce any “cultural shift” but rather transfers a well-known procedure from the individual to the collective (and national) level.

- Finally, although there was a strong political will to relate budget allocation to the evaluation results (funding criteria by the Ministry of University specify that 30% of the funding which is not increased with respect to the budget of the previous year “automatically” should be allocated on the basis of the result of scientific research activities), there is still little evidence of such consequences. Consequently, the competitive Italian RES can be characterized as “strong in principle”, but probably “weak in practice” as it is not sure whether the evaluation of research performance is consequential for funding.

**Introduction to VTR outcome**

**Source:** [http://vtr2006.cineca.it/ Final Report](http://vtr2006.cineca.it/)

For some time now and in many countries, strategies, procedures and instruments to evaluate research are object of public interest, and their application is normal practice. The Ministerial Order n° 2206 on December 16th, 2003 regulated the first Exercise of Evaluation of Italian Research for the years 2001-2003 (2001-2003 VTR). The Exercise was entrusted to the Italian National Committee for Research Evaluation (CIVR), and was aimed at evaluating the scientific performance of 77 universities (both State and non State), 12 State research institutions and 13 other research institutions (these latter participated willingly).

The present publication results from hard work that was conceived and carried out in record time. It is based on intuitions and new working methods, and the development
of the entire process of research evaluation represents a completely new experience for Italy.

The evaluation process

The 2001-2003 VTR has directly involved the 102 participating Structures. In fact, they were asked to autonomously select a pre-set number of research products outputted during 2001-2003 (books and their chapters, including congress proceedings; journal articles; registered patents and other research upgrading and transfer activities; projects, compositions, drawings and design; performance, shows and exhibitions; manufactures and art operas. Instead, purely editorial activities, books and software for solely teaching purposes, congress abstracts, tests and routine analyses, internal technical reports were not considered for evaluation). Structures were also asked to submit the above products on-line (considering the specificity of each of them) to one of the 20 Area Panels, which were the 14 standard scientific Areas of the Italian National University Council (CUN), together with 6 Special Areas, indicated by the CIVR.

Each Structure was asked to submit products for evaluation based on the number of Full Time Equivalent (FTE) researchers of the Structure itself. For universities each researcher is expressed as 0.5 FTE, since they have to fulfil two main duties (research and teaching). For research institutions each researcher corresponds to 1 FTE, having to fulfill only one duty (research). For this first Evaluation Exercise and in accordance with the Italian Scientific Community, the number of products was limited to 50% of the number of FTEs in each participating Structure. The products selected were 18,508, including about 1,178 of identical products submitted simultaneously by more than one Structure, and only one product was not evaluated because it did not correspond to the typologies required. Therefore, the evaluated products were actually 17,329.

Area Panels evaluated the products that were selected by the Structures. Each Panel was made up of 5 - 17 Panelists, according to the complexity of the specific Area and the number of products submitted. Globally, 151 Panelists were involved. The criteria for Panelist selection were set by the CIVR together with five Observers designated by the Scientific Community (Edoardo Boncinelli, Amilcare Collina, Umberto Colombo, Claudio Giovine, Daniele Marini). Panels autonomously chose 6,661 Experts who were asked to express a judgment of merit on the products. Each product was evaluated by at least two Experts according to the following criteria: quality, importance, originality/innovation and internationalization (for patents and research upgrading and transfer activities also the social-economic results, even potential, were evaluated). 59% of the Experts were from Italian Universities, 22% from foreign Institutions, 17% from Italian Research Institutions and 2% from Companies.

A total of 35,440 evaluations were carried out on the products selected (about 5 on average for each Expert). For each of the four criteria the Experts gave a written judgment on the product, but the rating of each product was unique and comprehensive, and each product was assigned with one of the following merit levels: excellent, good, acceptable, limited (excellent: the product is in the top 20% of the value scale shared by the international scientific community; good: the product is in the 60% - 80% range; acceptable: the product is in the 40% - 60% range; limited:
the product is in the low 40%). The Panels were asked to write a Final report (to be delivered to the CIVR) made up of three distinct parts. Consensus Report: based on the critical analysis of the judgments provided by the external Experts, in order to achieve a unique final judgment; Area Ranking list: which assigned the Structures with a merit score that was previously defined by the CIVR and calculated according to a weight assigned to each of the four judgment levels (excellent: 1; good: 0.8; acceptable: 0.6; limited: 0.2); Area final report: aimed at identifying the points of weakness and strength of the Area, with a specific section for patents.

Structures also provided the CIVR with data (international mobility; PhDs and postdocs; funds to be destined to research projects from the Ministry of Research and University - MIUR, from the European Union - EU and other international bodies, from other subjects and from own resources) to be correlated with product evaluation so to reach a final judgment on the Structure. Based on the information received, the CIVR developed a series of indicators related to:
- international mobility;
- training of young researchers;
- access to research projects by the MIUR;
- access to research projects by the EU and other international bodies;
- resources derived from other subjects and to be destined to research projects;
- resources derived from own balance to be destined to funding and co-funding of research projects.

Finally, Structures were asked to provide information on research upgrading and transfer activities:
- number of both patents registered during the three years and those active on 12/31/2003. Patents had to be separated by type: either registered in Italy or outside of Italy;
- revenues and costs during the three years derived respectively from patent selling and licensing and patent registration and management costs;
- descriptive and financial synthesis (in terms of receipts and costs) on all other transfer activities, particularly spin-offs and partnerships (only partnerships that have yielded the Structure total receipts > 500,000 Euros).

**Criteria suggested for resource allocation**

The criteria proposed by the CIVR for allocation of funds to the single Structure in relation to the outcome of the Evaluation Exercise were stated in the Call for Tender - VTR 2001-2003 (March 16th, 2004). Such criteria are based on the measurement of six parameters, each assigned with a specific weight. The CIVR has subsequently (July 4th, 2006) further explained how such parameters were determined and applied.

*Indicator A* - quality of the selected products (weight 4/9): for each Structure the share is calculated multiplying the rating of the selected products times their total number. The value is expressed as a percentage of the total Area value.

*Indicator B* - degree of ownership of the selected products rated as "excellent" (weight 2/9): for each Structure the share is calculated multiplying the degree of
ownership of the selected products rated as "excellent" times their total number. The value is expressed as a percentage of the total Area value.

**Indicator C** - international mobility (weight 1/9): for each Structure the share is measured in terms of (person years) researchers' in-/out-mobility (only periods greater than 3 consecutive months were considered and included in the final calculation). The value is expressed as the percentage of the total Area value.

**Indicator D** - post-graduate education (weight 0.5/9): for each Structure the share is given by the number of training researchers (PhDs and post-docs). The value is expressed as the percentage of the total Area value.

**Indicator E** - capacity to attract resources (weight 1/9): for each Structure the share is calculated by the funds for research projects derived from the MIUR, from the EU (and other international bodies), and from other subjects, with identical weight. The value is expressed as the percentage of the total Area value.

**Indicator F** - capacity to use own resources for research projects (weight 0.5/9): for each Structure the share is calculated by the funds for research projects derived from Structure own untied funds. The value is expressed as the percentage of the total Area value.

The share of the funds to be destined to each of the 20 Areas of evaluation (Q) was calculated multiplying the number of products submitted in the Areas times an Area coefficient (Area weight), in relation to the different costs of research in the various Areas. Area weights were determined through the investments for research projects derived from university own untied resources, referred to each product submitted in the specific Area.

Finally, another indicator was taken into consideration: **Indicator** - research upgrading and transfer activities: for each Structure the share is calculated by the number of patents registered during the three years (weight 1.5 for patents registered outside Italy), the number of patents that were active on 12/31/2003 (weight 1.5 for patents registered outside Italy), by the revenues from patent selling and licensing, by the number of spin-offs activated during the three years and by the number of partnerships (with receipts for the Structure >500,000 Euros), which were active for 2001-2003. The weights applied to the five items just mentioned were 1:1:2:4:2 respectively. The value is expressed as the percentage of the total Italian value.
4. Turkey: the limited institutionalization of the research evaluation system

As a foreword, one should mention here that little has been written on the Turkish research evaluation system, whereas issues pertaining to the evaluation of the quality of teaching have been more studied. The fact that information remains sketchy certainly pinpoints that the institutionalization of a Turkish research evaluation system is still very limited.

4.1. Research evaluation activities at the organisational level

**Public universities** Mizikaci’s monograph on higher education in Turkey (Mizikaci, 2006) mentions that the expansion of research activities within Turkish public universities suffers from several limitations. First of all, their main function remains teaching rather than research, although the higher education sector represents 60 percent of Turkey’s total research and development potential. Moreover, the amounts of state funding for research and development are low, and research projects often fail to receive financial support either from their institution or from the state.

Mizikaci thus states that research activities undertaken in higher education institutions are “largely disorganised in nature” so that one should not be surprised that the evaluation of research programs and of research organizations is still underdeveloped or even missing. In the same vein, the MEHEM country report on Turkey adds that many universities have not yet developed a research policy, and in such cases the unit that would coordinate research at the university level is missing.

However there has been a recent growth of project-based funding which has implied the expansion of ex-ante peer-review project evaluation. The most prestigious grants are allocated by the Scientific and Technological Research Council of Turkey (TUBITAK, see [http://www.tubitak.gov.tr/](http://www.tubitak.gov.tr/) and MEHEM country report on Turkey).

Moreover the few existing procedures for evaluating research activities show that there are some limited attempts to introduce some features of both “strong and competitive” and “weak and intrusive” RES.

“**Strong and competitive**: the Interuniversity Council⁵ appoints an evaluation committee in order to rank the state universities according to the average number of articles published per teaching staff member in prominent academic journals. The Council of Higher Education (YOK) could choose to take these reports into consideration and to make the ranking consequential for funding during the budget negotiations with the public universities and the ministry of finance. However, there is no evidence that these rankings have a concrete impact on the bargaining process and finally on the budget allocation.

“**Weak and intrusive**: the Faculty Performance Reports submitted by each faculty member at the end of each academic year about its teaching and research

---

⁵ Which is comprised of all the rectors of all universities and one member elected by the senate of each university.
activities are aggregated into indicators at the university level, and then submitted to the Council of Higher Education (YÖK). YÖK then makes recommendations in order for universities to improve the quality of their teaching programs or their research performance. Here again, there seems to be little evidence about the scope and the actual consequences of such recommendations.

**Private universities** Traditionally, most of them focus upon teaching more than they do upon research, with only a few large institutions active in establishing research centres and cooperative, international projects. However, in order to be eligible for the State financial assistance, private universities have to fill a list of criteria pertaining to their teaching as well as to their research activities. The amount and share of state subsidy is quite limited (less than 4% in 2005). It is related to certain indicators of performance of the university.

- have provided formal education for two years,
- have a publication /teaching staff member ratio equivalent to that of that of those state universities occupying a position in the upper half of a ranking of state universities in terms of average number of articles published per teaching staff member in a prominent academic journals recognized by an evaluation committee appointed by the Interuniversity Council,
- be in a position equivalent to those state universities in the upper half of a ranking based on the number of students placed there who ranked among the top two thousand in terms of science and mathematics scores on the student placement examination,
- grant full-tuition scholarships to a minimum of fifteen percent of its students (CHE Website, 2009)

### 4.2. Faculty performance evaluation

**Public universities.** Academic staff members in public universities are civil servants who are not regularly evaluated apart from those who apply for a promotion as associate or as full professor. The criteria for promotion include research activities, but also teaching and administrative responsibilities. Consequently, and given the fact that faculty members have over-extended teaching loads, there is little incentive for faculty members to improve their research performance at Turkish public universities.

Under these conditions, Mizikaci (op. cit.) even argues that any evaluation procedure would suffer from a general distrust, as it “would not result in any career promotions or reductions”.

Consequently, the Faculty Performance Report on teaching and research that each faculty member has to submit at the end of each academic year seems to be inconsequential.

**Private universities.** Private universities are the employing organization of their academic staff. Faculty performance evaluation is much more developed than in public universities because private higher education institutions compete against each other to attract students. Therefore most of them have established Faculty performance evaluations on research and teaching activities, according to their own
criteria. These evaluation procedures can be characterized here as “strong and intrusive” as they rely on qualitative, in-depth information and as they have consequences in terms of career progression and yearly salary raises.
5. France

As a MEHEM report has been dedicated to the analysis for the French evaluation system (Louvel and Gonthier, 2009b) the present report will synthesize its conclusions and will relate them to the theoretical framework on which the international comparison relies here.

5.1. The evaluation of public research institutions: from “weak and intrusive” to “strong and intrusive” and/or “strong and competitive”?

A weak and intrusive research evaluation system until 2006. Mixed research units⁶ were evaluated by disciplinary evaluation committees from the public research organization (PRO) they are associated with. Evaluations were mostly based on peer reviews: they took into account the scientific production, as well as how the research units’ projects fit into the global strategy of the PRO, and the quality of the internal management. Mixed research units were evaluated at least every four years. University research units which were not affiliated to any PRO were evaluated by an administrative body under the authority of the Minister for Higher Education and research (“Mission Scientifique Technique et Pédagogique”).

The evaluation procedures were clearly intrusive as they relied on qualitative, in-depth information which was not synthesised in order to compare the research units under assessment, but was rather used for internal purposes. Moreover, the research evaluation system was rather weak as it did not have systematic consequences on funding (although one should not be too simplistic: poorly performing research units could suffer from budget cuts or be denied position openings).

The creation of the higher education and research evaluation agency (AERES) in 2006. The AERES now evaluates all research units according to the same set of criteria which explicitly refers to international standards. The evaluation relies on a one-day or two-day site visit by an evaluation committee⁷ and it follows the guidelines and criteria provided by the AERES. The committee then produces a report and transmits it to the head of the AERES department for the evaluation of higher education and research institutions.

There are still some uncertainties whether the new research evaluation system can be considered as “strong and intrusive” or as strong and competitive”. In deed, there a clear political will to link the output of the evaluations to the budget allocation to the research units. However, it is not clear whether this link between evaluation and funding will be established in an “intrusive” perspective, that is if the evaluation

⁷ The evaluation committee is made up of experts chosen among a list of 4000 academics, and of experts put forward by the evaluation committees of the PRO or universities the research unit is affiliated to.
procedure will mostly serve organizational purposes of managing research, or if it will also foster competition between research units. What causes this uncertainty is not so much the use of quantitative indicators (which is highly debated among the French scientific community) than the fact that AERES assigns grades\(^8\) to the research units and publishes them on its website.

### 5.2. The evaluation of individual academics: still “weak and intrusive, but more consequences on career progressions”

Since April 2009\(^9\), all French academics (university academics as well as PRO researchers) are evaluated every four years. Their evaluation is conducted by disciplinary committees (from the National University Board – CNU – in the first case, from the PRO they are affiliated to in the second). The evaluation mostly relies on peer review. Since 2006, the AERES has to approve the evaluation procedures and also provides some guidelines, such as disciplinary lists of journals where academics should publish in order to be considered as “actively publishing researchers”. The AERES presents these guidelines as a way to create to harmonise the evaluation of individual academics and that of research organisations which falls under its mission.

The evaluation of individual academics can be considered as “weak” insofar as it has no consequences on their salaries. However, evaluations of university academics have now an impact on the determination of their teaching duties. In deed, the university president (or the head of the higher education institution) can since 2009 modulate them with regards to their research activities as evaluated by the CNU – however with the consent of the university academics. In that sense the new evaluation procedures have greater consequences on the university academics’ ability to conduct research and then on their career perspectives.

Moreover, the evaluation is still “intrusive”, as it is mainly used by universities and PRO to improve their own research performance. However, as in the Spanish case, one can expect some limited shifts to a “competitive” evaluation system, which would be related to the use of data collected on individual researchers in an aggregated form performance. In deed, university departments and PRO could for example use the number or the percentage of “actively publishing researchers” among their academic staff in order to improve their collective reputation in the context of an increased competition between research organisations. However there is still no clear indication of this “competitive” use of the new evaluation procedures.

---

\(^8\) The grades are not attributed by the assessment panels, but by disciplinary “grading committees” (Comités de notation) composed of presidents of assessment panels and of the AERES scientific delegate. The grading committees’ mission is to examine all evaluation reports, to “harmonize” them and to attribute grades on the basis on this comparative evaluation. The grading procedure has been recently modified in order to increase the relevance of the grades attributed. When the first evaluations were carried out in 2007, research units and teaching curricula were attributed a single grade (A+, A, B, C), a procedure which was meant to maximize the comparability of the evaluation outcomes. However, researchers themselves and policy-makers had difficulties interpreting this unique grade, so that the outcomes of further evaluations were several grades (A+, A, B, C), still synthesised at the end by a single grade.

\(^9\) Promulgation of a decree modifying the status of French university academics, 25\(^{th}\) of April 2009.

University academics were previously only evaluated when they applied for a promotion as full professor.
6. Synthesis

<table>
<thead>
<tr>
<th>Spain</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Since 1989</strong></td>
</tr>
<tr>
<td><em>Evaluation of individual researchers</em></td>
</tr>
<tr>
<td>Intrusive: qualitative information, gathered through peer-review. Information is used to increase the productivity of individual researchers (but also in From intrusive to competitive? The use of data from individual research performance in aggregated form → comparison of collective reputations; national publications similar to rankings (since 2004)</td>
</tr>
<tr>
<td><strong>Since 2001</strong></td>
</tr>
<tr>
<td><em>Evaluation of individual researchers</em></td>
</tr>
<tr>
<td>Intrusive and weak: accreditation and habilitation national procedures in order to get access to resp. pre-tenured and tenured positions.</td>
</tr>
<tr>
<td><strong>Since 1987</strong></td>
</tr>
<tr>
<td><em>Evaluation of research organizations</em></td>
</tr>
<tr>
<td>Intrusive and weak</td>
</tr>
<tr>
<td>Marginally developed compared to the evaluation of individual researchers and of the ex-ante peer-review evaluation of research projects.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Italy</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Since 2001</strong></td>
</tr>
<tr>
<td><em>Evaluation of public universities and public research institutions</em></td>
</tr>
<tr>
<td>VTR: Disciplinary-based evaluation</td>
</tr>
<tr>
<td>Competitive (qualitative information, peer review) — evaluations are synthesized in a single-scale rating which is considered as an easy way to compare the performance of the research entities</td>
</tr>
<tr>
<td>Strong in principle (a strong political will to relate budget allocation to the results of the evaluation), but weak in practice?</td>
</tr>
</tbody>
</table>

| Turkey                                                               |
|                                                                     |
| **Limited institutionalization at the organizational level**         |
| Some attempts to organize weak and intrusive as well as strong and competitive evaluation procedures |
| **Evaluation of individual researchers**                             |
| - Public universities: inconsequential Faculty Performance Reports |
| - Private universities: strong and intrusive (consequential for salaries and career progression) |

<p>| France                                                               |
|                                                                     |
| <strong>Evaluation of public research institutions</strong>                       |
| Since 2006: from “weak and intrusive” to “strong and intrusive”      |
| (consequential for funding; used for purposes of internal management only) and/or competitive? (consequential for funding; publicly available grades enabling comparison and competition between research units) |</p>
<table>
<thead>
<tr>
<th>Evaluation of individual researchers</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Weak</strong> no consequences on salaries, but can be consequential for the annual teaching duties of university academics (which are modulated according to whether they are considered as &quot;actively publishing researchers&quot;).</td>
</tr>
<tr>
<td><strong>Intrusive</strong> information is not publicly available and used for the management of researchers only. However, as in the Spanish case, it can have some consequences on the establishment of collective reputations.</td>
</tr>
</tbody>
</table>
7. References


