Research Evaluation and Governance: a Comparative Approach

Jordi Molas-Gallart
INGENIO (CSIC-UPV), Universitat Politècnica de València, Camí de Vera s/n, 46022 València, Spain.
jormoga@ingenio.upv.es

Abstract
Through a comparative study of the UK and Spain, this paper addresses the effect of different governance systems and administrative practices on the functioning and role of research evaluation. It distinguishes three main evaluation functions: learning, distributive and accountability. In the UK, a flexible research management structure can respond to evaluation outcomes, and the three functions of evaluation co-exist in a diversified, decentralized, evaluation system. Despite rhetorical similarities, the Spanish evaluation system is dominated by its accountability functions, often playing the role of an audit system, and is mainly the responsibility of specialized evaluation agencies. These differences affect the way in which “evaluation” is understood in both countries. They cannot be attributed to a differential development in “evaluation cultures”, but rather to different research governance systems affecting the nature and scope of evaluation practice.

Introduction
The role of evaluation within the policy process depends on the administrative system within which the evaluation practices are inserted. Although this point has long been recognized its implications for evaluation policy learning and practice are often overlooked. In the late 1980s a comparative study observed that the ways in which evaluation was approached in different countries “reflected” their political and administrative culture (Gibbons & Georghiou, 1987). Reporting on another comparative set of studies Georghiou attributed the differences to three different factors: (1) the state of development of the research infrastructure, (2) the ways in which science is organized, and (3) the general practices of governance beyond the research domain (Georghiou, 1995, p. 4). In the study on Spain that constituted part of this comparative effort, Sanz-Menéndez argued that “the evolution of research evaluation activities or practices could be viewed as embedded in the institution for governance of the R&D system and in the general characteristics of the system for making public policy” (Sanz-Menéndez, 1995, p. 80). The ways in which different governments have tried to manage public science through the introduction of different evaluation systems, and how the resulting governance structures affect scientific production have been the object of some research (Whitley & Glässer, 2007).
Yet, most evaluation literature focuses on discussing the different approaches to evaluation on their own merits, as if they were, in practice, independent from the administrative context where evaluation takes place. The debates about the functions of evaluation and the degree to which concepts like “formative” and “summative” provide a useful classification of evaluation types seem to assume implicitly that evaluation is in the driving seat; that is, that the form of evaluation we adopt will define the nature of the policy process in which evaluation is inserted. Furthermore, it is at times specifically argued that countries progress through different evaluation culture stages, evolving towards increasingly sophisticated evaluation practice (Toulemonde, 2000). This approach carries with it a simple “policy learning” message: countries with comparatively lesser experience of evaluation should follow the practices of the countries that enjoy a more developed evaluation culture.

This message is often translated into practice: countries that are relatively newcomers to the field of evaluation are importing evaluation methodologies, and their accompanying foreign experts and consultants to help them develop and implement evaluation strategies. Little consideration seems to be paid to the political and administrative framework within which evaluation takes place, and to the extent to which differences in political administration influence the practice of evaluation. It is stated that countries like Spain are lagging in its evaluation culture, and such lag is attributed to the dearth of evaluation experience, the lack of formal training in evaluation for professionals and civil servants, and the lack of formally established evaluation standards (Bustelo, 2006).

This paper reconnects with the line of comparative research that more than a decade ago linked research evaluation practice with the broader research governance structure. The paper first reviews how the literature has discussed the different types and functions of evaluation, and settles on three different functions of evaluation that we will use to compare research evaluation practice. It then shows how academic research evaluation practice in the UK and Spain focus on different functions; in fact, although the term “evaluation” is used in all contexts it refers to very different activities. The final section discusses how these activities fit within the different systems of research governance. It is therefore misleading to speak of an “evaluation culture” as if this was an independent variable that we can develop without reference to broader governance issues.

Evaluation and its functions

---

1 By academic research I refer to research carried out mainly in Higher Education Institutions and other Public Sector Research Establishments with the main objective of expanding the knowledge base.
As concepts gain popularity they often lose precision. “Evaluation” is no stranger to this problem: the concept can be used to refer to very different types of practice. A standard dictionary definition will equate evaluation with “valuation”. The American Heritage Dictionary, for instance, defines evaluation as ascertaining the value of something. Scriven, one of the doyens of evaluation theory, has defined evaluation as “the process of determining the merit, worth and value of things, and evaluations are the product of that process” (Scriven, 1991, p. 1). But this is extremely broad. It includes both formal, structured inquiry and informal assessment without the support of explicit criteria. Neither does this definition address the objective of evaluation. Evaluation analysts have narrowed the definition to apply it specifically to the task of professional evaluators; that is, the evaluation of public policy interventions. Rich defined evaluation as “the process of assessing whether or not desired or undesired outcomes have been reached, of specifying or explaining the outcomes that were reached, and of suggesting new strategies and/or definitions of future problems” (Rich, 1979, p. 10). There are two aspects to note in this definition. First it implicitly refers to specific policy interventions. Second, it introduces the notion that one of the goals of evaluation is to contribute to future policy formation by, for instance, “suggesting new strategies”. Further, in the same book, Rich argues that a critical function of evaluation is to contribute to organizational learning (Rich, 1979, p. 80). These traits are further reinforced by Chen’s notion of “programme evaluation” as “the application of evaluation approaches, techniques, and knowledge to systematically assess and improve the planning, implementation, and effectiveness of programs” (Chen, 2005, p. 3). The focus is firmly placed on specific types of policy interventions (“programs”) and the goals of evaluation are defined as both assessment and improvement.

Chen and Rich’s approaches coincide with those of many evaluation practitioners in placing the objective of policy learning squarely at the centre of what evaluations are supposed to be for. Yet, there are other traditional functions of evaluation.

Evaluation can be conducted to ensure the accountability of those using public resources to provide goods and services to society. Evaluation for accountability will provide “an external assessment of the effectiveness, efficiency, value for money, and performance” of a policy or set of policies (Batterbury, 2006, p. 182). Accountability-oriented evaluations are a tool of democratic scrutiny over the organizations and individuals using public funds to implement public policies.

Evaluation can also be carried out as a tool to help distribute resources or rewards among policy implementers and beneficiaries. In such “distributive function” (Cruz-Castro & Sanz-Menéndez, 2008) the allocation of resources is decided according to the merit that the evaluation attributes to different individuals. Ex-ante project evaluation would be one example of such distributive type of evaluation, another being the use of evaluations to distribute rewards to individuals or groups that have performed according to some pre-established criteria.
This paper will therefore distinguish between the learning, accountability and distributive functions of evaluation,\(^2\) arguing that, although it is possible for different functions to be combined within a single evaluation, in practice the different functions translate into different approaches to evaluation and its organization. I will illustrate this by comparing the British and Spanish approaches to academic research evaluation.

Research evaluation in a comparative perspective

Research Evaluation in the UK: a multifunctional perspective

We can distinguish two main sources of research funding: (1) core funding supporting academic research for the long-term and granting academics high degree of freedom in the selection of research topics, and (2) project funding for clearly defined, time-bound specific research initiatives. In the UK, core funding of research activities is organized through a formula-based approach that allocates money to universities according to their past research performance. The formula is based on the rating that the university departments achieve in the “Research Assessment Exercise” (RAE) (now being replaced by the Research Evaluation Framework -REF) and the number of academic staff involved in the assessment.\(^3\) Therefore, core funding is organized at the institutional level and is only assured for a period of years until the following evaluation exercise. The RAE (and now the REF) are the core evaluation activities in this process: and assessment of past research performance conducted for distributive purposes. The subjects of these evaluations are university departments through the RAE/REF system, and the process is managed by the Higher Education Funding Councils of the different British regions, who are in charge of defining and implementing the funding instruments. The specific assessments are commissioned to panels of academics, and occasionally other experts, who review the evidence presented by the University departments and their scientific production. The panels agree a rating for each of the units under review following a set of assessment criteria. The criteria, submissions and panel decisions are publicly available. This leads to a process that is both extensive (as it covers the whole of the English Higher Education system) and intensive (as each department’s report and the outputs it presents needs to be assessed, individually, by the assigned panel). The approach does not deploy an indicator-based system of formal measurement, but so far has required a very large investment of resources.

\(^2\) It must be noted that authors have identified other functions like legitimization and providing a forum for policy debate (Luukkonen, 2002, p. 84), or have used different terms and slightly different conceptualizations to refer to similar distinctions. See, for instance, Chen (1996) for a typology and an explicit critique of the most commonly used differentiation between formative and summative evaluations.

\(^3\) For detailed analysis of the RAE operation, consequences and debates on their evolution see Barker (2007) and Martin & Whitley (2010).
Project funding is managed mainly by the Research Councils, who use a variety of instruments, from individual doctoral grants to multi-million, multi-year research centers. A large share of these grants is used to fund personnel costs: council-funded researchers contracted to carry out specific research projects are an important component of the British academic system. In this case again, the evaluation processes are controlled by the same organizations that are in charge of policy implementation; there are no specialized evaluation agencies.

Ex-ante project appraisal and ex-post assessment of final reports is carried out through a peer review system organized by the Councils themselves. The reviewers’ comments tend to be detailed and are distributed to the applicants; some Councils allow applicants to respond with comments before a funding decision is made.

In addition ex-post impact assessment are very important for learning and accountability purposes. Substantial ex-post evaluations focusing on the impacts of specific investments are often carried out under contract by specialist consultants. Research Councils UK (RCUK), an organization that brings together the main UK academic research funding organizations, has a “Performance and Evaluation Group” in charge of “providing strategic direction on all issues relating to evaluation and benchmarking including the evaluation of Science Budget investments in research, training, knowledge transfer, science and society activities and operational performance” (http://www.rcuk.ac.uk/aboutrcuk/executivegroup/subgroups/pegroup.htm). Among other objectives the group seeks to coordinate the evaluation activities of the different Research Councils and share best practices. Within the Research Councils different groups are in charge of different evaluation tasks; thus, for instance, in the Economic and Social Research Council (ESRC), ex-post evaluation is organized by the Research Evaluation Committee (REC). The REC commissions ex-post evaluations of its programs (targeted funding, for a period of a 5 or 10-year, of a number of interrelated research projects), centers (funding of a substantial group of researchers working on the same field over a period of 5 or 10 years), and independent research projects. The Research Councils have paid considerable attention to the development of evaluation research methodologies, which are generally based on building a detailed understanding of the processes through which impact takes place. For instance, a far from exhaustive review of reports and papers commissioned by a single Research Council (ESRC) yields well over a dozen of publications and reports that are either methodological reflections or evaluations that include novel methodological development as part of their remit (Caswill, 1994; Cave & Hanney, 1996; Davies, Nutley, & Walter, 2005; Faulkner, 1995; LaFollette, 1995; J. Molas-Gallart & Tang, 2007; J. Molas-Gallart, Tang, P., Sinclair, T., Morrow, S., Martin, B., 1999; Nutley, 2005; Redclift & Shore, 1995; 4 Similarly the Engineering and Physical Sciences Research Council has a “Performance and Evaluation Team” operating along similar lines to REC.

5 Naturally each specific project, program or centre is, in addition, subjected to peer review assessment at the proposal stage and, once completed, to assess the final report.
Rip & van der Meulen, 1995; Tang, Sinclair, Holmes, Wallace, & Hobday, 1998; Tuck, 1995; Whiston, 1990; Wooding, et al., 2007). The results of these evaluations are publicly available.

In short, evaluation of academic research investments is devolved to the organizations in charge of policy implementation and is carried out in a decentralized manner: mainly by academic peers for project appraisals and often by independent paid consultants for ex-post impact assessment. As a consequence, a competitive evaluation marketplace has evolved with a number of consultancy companies and university groups and departments actively providing evaluation services to the Councils.\(^6\)

The UK research evaluation system plays, through its different processes and tools, all three main evaluation functions. Evaluation processes directly linked to the management of research resources have mainly distributive functions. Ex-post evaluations to assess the impacts of research investments and the process through which such impacts take place have policy learning and accountability functions; the ex-post impact assessments carried out by the Research Councils seek to acquire information on impact processes and to use this information to inform the design of research support and exploitation tools. Inasmuch as they identify social value attributable to the research investments, they play an accountability role as well.

The learning function is also present in the traditional assessment routines associated with research management. Although the peer review of project proposals is part and parcel or the normal administration of research funding organizations, in the UK peer reviews tend to be detailed assessments of the proposals and they are always distributed to the researchers. Although their primary function is to support decisions on the allocation of funds, the researchers can use the information they receive from the assessments to derive lessons for future proposals.

Research evaluation in Spain: the persisting dominance of the accountability function

The Spanish research system is characterized by the prominence of core funding channeled through the salaries of individual tenured academics. A percentage of the working time of tenured lecturers is assumed to be invested in research (and therefore accounted in the official R&D statistics as government R&D expenditure). In addition, the central Government funds several Public Research Establishments, the most important of which, the Spanish Council for Scientific Research (CSIC), \(^7\) employs some 2200 full time researchers organized in a plethora of research groups and institutes. Core funding for these institutes is

\(^{6}\) Similar characteristics have been identified by Feller (2003) for the US system.

\(^{7}\) Unlike the homonymous British “Research Councils”, CSIC is a research performing government organization until recently dependent from the Ministry of Education and recently converted into an Agency. As a research organization CSIC is closer to the French CNRS or the Italian CNR.
composed of the salaries of their tenured staff and a related overhead component.

Spanish research project funding revolves around the National R&D Plan. This is managed by a single agency, the National Agency for Evaluation and Prospective (ANEP). National Plan funding is distributed among a large number of university and CSIC groups and tends to fund marginal costs and junior doctoral grants associated with research projects: it is very rare for a National Plan project to fund the salaries of senior researchers, who are almost always tenured academics within the Spanish system.\(^8\)

In contrast with the United Kingdom, Spanish research evaluation is managed through specialized evaluation agencies. The National Agency for Evaluation and Prospective (ANEP) is in charge of organizing the peer review evaluation of research proposals submitted to the National Plan and the review of the projects’ interim and final reports. It therefore focuses on the project component of the funding system and plays a purely distributive function. The system is shaped by the large throughput of proposals and individual assessments that need to be dealt with a “weak bureaucracy” (Cruz-Castro & Sanz-Menéndez, 2008) managing very limited resources given the size of the task at hand.\(^9\) Their evaluations are necessarily cursory and the reviewers comments are not forwarded to the applicants: the interaction between applicants, managing agency and peer reviewers is kept at a minimum. The decisions and assessments that are conveyed to applicant and project-holders are very succinct,\(^10\) and no correspondence will typically be entered between researcher and reviewer. The reviewing process provides very little information to the applicants as to how to improve the project in the future and, therefore, does not play a learning function.

Another agency, the “National Commission for the Evaluation of Research Activity” (better known by its Spanish acronym CNEAI) is in charge of implementing a system by which all Spanish tenured academics can submit, every six years, evidence of the results of their research activity. If the Commission deems that the proof submitted (crucially a list of the five most relevant publications of the period) constitute evidence that the individual has been research active, it awards a “sexenio”: an official confirmation of research activity that carries with it a modest increase in salary for the rest of her life and into her pensions (which in Spain is earnings-related). The main role of the Commission is to carry out this assessment and to this end it draws on a rotating

---

\(^8\) In addition to central government funding, many Spanish regions have launched their own plans in support of research and innovation activities. These are managed by regional agencies and vary in organization, objectives and funding from region to region. In general, however, they fund research projects along similar lines and conditions to those implemented through the National Plan.

\(^9\) ANEP is led by seconded academics and has a very lean core organization.

\(^10\) For instance, referees’ comments on proposals are not conveyed to the applicants.
staff of academics appointed for a fix period to these functions. The process by which “sexenios” are granted and the role played by CNEAI in this processes, emerges as the main mechanism for the ex-post evaluation of Spanish tenured researchers; in other words, it is the main evaluation tool for the core component of Spanish research funding.

The resources at CNEAI’s disposal to assess thousands of academics every year are, again, very limited. Consequently, its decisions are driven by sets of “indicators” that are narrower than the official assessment criteria. The CNEAI assessment system focuses, in practice, on the number of journal articles published in ISI-listed journals and other academic publications preferably in foreign-language scholarly books and journals. There are thresholds that the individual aspiring to obtain a “sexenio” must pass and individual assessment of the quality of the publications is effectively precluded by the lack of resources. Finally, CNEAI communicates its decision to the individual in a terse manner: a phrase justifying the decision is accompanied by a long paragraph explaining how the academic can seek redress and the legal basis for this process (specifically 3 legal articles from two different pieces of legislation, stemming from two different ministries). The way in which the decision is communicated emphasizes the administrative nature of the process, which in practice is converted into an auditing process. The system has become an auditing mechanism to confirm the authenticity of the claims submitted by the academics; a system of incentives designed to play a distributive function has become a process to reinforce accountability from tenured academics.

Importantly, the CNEAI system focuses on individual researchers, and the ANEP focuses its attention on the appraisal of individual research proposals and the auditing of ongoing projects. The importance of individual evaluations, instead of organizational ones, has been identified as one of the cardinal traits of the Spanish system of evaluation (Cruz-Castro & Sanz-Menéndez, 2008). The evaluation of structured interventions (as, for instance, of a research “plan” funding a large number of projects) has been limited to some attempts by regional authorities, and the evaluation of CSIC research institutes and centers carried out under the CSIC programming process.

Since the mid 2000s, CSIC has rolled out another evaluation system: an ex-ante assessment of institute-level strategic research programs, which also takes into account past research performance. The whole organization develops every four years a strategic plan is informed by the plans that each formally recognized

---

11 Detailed descriptions and assessments of the “sexenios” system can be found in (2008); and (Jiménez-Contreras, de Moya Anegón, & López-Cózar, 2003)

12 Outside the central government agencies some regional governments have set up their own organizations to discharge similar functions within their own regions. In particular many regions have developed their own Science and Technology Plans, and other programs to support research activities. In some cases, project selection is contracted out to the ANEP, but some regions have built their own evaluation agencies, which in some cases are broadening their work beyond project evaluation to ex-post program evaluations and ex-ante policy definition studies.
CSIC research group or institute has to develop. These plans are assessed by committees of foreign peer reviewers that can suggest changes to the objectives and targets defined in the plans. Therefore, the evaluation appears, on the face of it, to play a learning role: plans are presented and discussed with the evaluation panel and the Institutes have to respond to the comments made by the reviewers. Formally, the agreed strategic plan sets objectives for the groups and establishes a plan for the transfer of resources (importantly tenured positions) to these groups to carry out their agreed strategies. In practice, however, the approved strategic plans do not amount to a commitment for any of the participating "partners" (central CSIC administration and research groups and institutes), and there is no established mechanism to follow the implementation of its recommendations. What remains, however, is a limited set of annual quantitative targets (publications, funding raised,...) against which the groups are measured every year and if deemed successful their employees receive a small productivity bonus. What had been designed as an evaluation strategy with a strong learning function has become another auditing system for accountability backed up by a weak distributive tool in the shape of an economic incentive.

The 2008 National R&D Plan made reference to the latest newcomer to the Spanish research evaluation system: the establishment of an “Integral Monitoring and Evaluation System” (SISE in its Spanish acronym). Originally an initiative launched by the Spanish Foundation for Science and Technology (FECYT) to monitor the National Plan and other research programs it is yet to be turned into an explicit structured approach, let alone an evaluation “tool”. Yet it is worth noting how the National Plan describes SISE’s objectives:

“The Integral Monitoring and Evaluation System (SISE) is the tool designed for management control of all public programs in support of R&D&I, to improve transparency and the publicity given to all interventions, so that the citizens and the Spanish society can be informed about the activities that are being supported with public funds” (my translation of the Spanish version available at www.plannacionalidi.es/gestion/seguimiento.php )

This paragraph is clear in determining the main function allocated to the fledgling system: the SISE is a monitoring tool to provide control and accountability of public actions. It is therefore apparent that the learning function is neither an objective in current evaluation practice, nor is it included among the goals for the development of an evaluation system.14

Evaluation and Governance

---

13 The FECYT is a foundation dependent from the Ministry of Science and Innovation to support the Spanish innovation system through R&D, collaboration and dissemination activities.

14 Some ex-post program level evaluations have been conducted by regional authorities assessing the implementation of their own regional research plans. These could in principle perform a learning function but so far this has been precluded by their weak linkages with the policy process. The evaluations have tended to be ad-hoc ex post exercises with mainly a learning function which in practice have been limited to the direct commissioners of the study (evaluations are seldom published), and without a established feedback system into the policy process.
The systems we have briefly reviewed in the preceding section are naturally complex. The UK and Spain are large countries with complex administrative structures implementing a variety of research policies. Evaluation strategies and practices can vary from actor to actor, and may combine different functions within a single evaluation strategy. Yet, the comparison of the UK and Spanish systems has exposed that they have little in common concerning the dominant practices and functions of research evaluation. Perhaps the most relevant difference is in terms of what the Spanish approach to evaluation does not do. Unlike the UK, research evaluations in Spain do not, in the main, deal with program-based evaluations (a difference referring to “what” is being evaluated), seem unconcerned with the analysis of the processes through which effects take place (a difference in “how” evaluations are conducted) and do not attend to the learning function of evaluation (“why” evaluation is done differs). While the British system combines the three functions and pays attention to the development of approaches to further evaluation’s learning functions, the Spanish system focuses on accountability and distributive functions and operates as a complex system of administrative controls.

How can we explain such differences? As we pointed out at the start of this paper, one can see the disparities as reflecting different stages in the development of evaluative practices. From this perspective, countries like the UK with longer experience in the implementation of policy evaluation have been able to deploy systems that allow its results to be fed back to stakeholders, thus developing its learning function. In comparison, the argument would continue, a country like Spain, where evaluation is a more recent practice, feedback processes and the learning function they support are yet to be developed. This argument is however based on a strong assumption: that evaluation practice can develop independently from administrative culture and that it is through changes in evaluation approaches that the administrative and overall governance systems can change. This is a tall order for a fledgling practice.

As we have discussed at the start of this paper, there is an alternative view that stresses the links between evaluation practice and the governance system, including the dominant administrative culture. Here the assumption is that evaluation approaches cannot develop independently; on the contrary, they are explained by the administrative processes within which they are inserted. The variance in evaluation approaches can therefore be explained by the existence of different governance systems.

An analysis of the two cases I have compared here suggests that the differences in research administration and governance can go a long way in explaining the diversity in evaluation practice. We can identify two key, interrelated, factors with a bearing on evaluation practice: the extent of tenured appointments within the academic system, and the role of project funding. Tenured appointments are not universal in the British academic system, often the reward to senior academics who have reached the rank of professor. Instead, it is common for even senior academics to derive a part of their salaries or their totality from research projects throughout extensive periods of their careers. The research project thus
becomes a crucial tool to sustain the individual activity of many academics and to define research priorities and activities. Having to support salaries as well as all the other research costs, UK research grants have tended to be financially large in relation to the staff formally involved, and become an important determinant of research activity. Consequently, research activity tends to be project-based. Also, bigger sized projects mean a less fragmented portfolio of research activities: a relatively (to the size of the budget) lower throughput of proposals and projects, which can play a determinant role in the careers of academics and the overall profile of research activities. In this context, peer review (both of research proposals and final reports) and ex-post program evaluation are called to play a learning as well as distributive and accountability functions. Detailed peer reports are important for the development of future projects, and the lessons of previous research initiatives identified through program evaluations can be translated into new priorities and funding mechanisms with a rapid effect on the overall research strategy of the academic community. The importance of projects makes the system more flexible and responsive, and therefore the lessons derived from evaluation can be conveyed into practice with relative ease.

The governance of academic research in Spain is very different from the situation I have just described for the United Kingdom. First, it is a system dominated by tenured appointments: most established academics are tenured public employees, working either for their public universities or for public research establishments like CSIC. Access to academic tenured jobs is regulated, as for any public servant, by standard bureaucratic controls administered centrally to ensure fair and equal access conditions. Further controls are established to implement a degree of accountability on the performance of the functions for which the scientists/public employees are responsible. These processes necessarily pose constraints to the ways in which academics can organize their activities. Hiring processes, in particular, are cumbersome and subject to multiple conditions. Consequently, the personnel structure of Spanish academia is very rigid. The system that emerges from these practices is difficult to adapt to the potential recommendations that could emerge from “formative” evaluation processes. Any recommendation that would point to a perceived necessity to incorporate new skills, or to change research priorities would be in practice difficult to implement, at least in the short term.

---

15 Particularly if compared to those in countries, like Spain or Italy, where the majority of academic personnel is tenured.

16 “Core” research funding allocated through the Research Assessment Exercise is an important source of funding that is not linked to any specific research project or strategy. Yet the amounts of funding that universities receive through this channel are revised every 5 to 7 years. Currently funding is linked to the results of the Research Assessment Exercise, and a formula-based system is being developed for use in the future.

17 There are Spanish private universities but their relative importance is small except in some specific disciplines like business administration.
Further, within this structure, the role of the research project is very different from the function that the same tool plays in the British case. Projects like those funded by the Spanish National R&D Plan and other academic research programs implemented by regional authorities and other organizations will only cover “marginal costs”: those additional costs attributable to the research activity. Direct salary costs are already covered by existing budgetary allocations (they are part of the “core funding” of Spanish research activities). Consequently, research projects will tend to be relatively small in financial terms,\(^\text{18}\) and, more importantly, as in many disciplines the majority of research resources are financed through “core funding” streams it is unlikely that academics will organize their activities on a project basis. In this situation it is more difficult to influence research agendas through changes in project funding programs.

Altogether, the relative minor role of projects and the tenured status of research scientists combine to constitute a system that is not responsive to attempts to “fine tune” its management or to steer its priorities. The learning function of evaluation cannot be as effective as in a more flexible governance structure. Instead, the bureaucratic nature of the research management system calls for auditing and accountability mechanisms. It is not surprising that these are dominant in the Spanish approach to evaluation.

This paper has therefore confirmed the insights of analysts that more than a decade ago were suggesting that national approaches to evaluation reflect local administrative culture and practice. By comparing two contrasting cases I have characterized these differences in terms, mainly, of the different functions that evaluation plays in different contexts. There is a clear policy implication that can be derived from this observation: to draw policy lessons on evaluation strategies from countries with different research governance structures will be ill-advised, unless these lessons are integrated within a broader strategy of administrative reform.

Finally, and from a practical perspective, the differences we have identified are so profound that they lead to a different understanding of what “evaluation” means. The dominant notions of evaluation in the British context are aligned with current program evaluation practices emphasizing their learning function. In Spain, evaluation is dominantly seen as part of a broader system to ensure accountability and manage a resource allocation system; often no more than an administrative task. The increasingly common rhetoric with its calls to “strengthen evaluation” is likely to be interpreted very differently by communities of research policy practitioners in different countries. The potential for confusion lies in the fact that we may be assuming (incorrectly) that we are all speaking the same evaluation language.

\(^{18}\) This situation varies across research fields. In areas like the social sciences where the need for laboratories and other research infrastructures is often small, successful projects will receive small contributions (for travel, empirical research costs, doctoral students assigned to the project). Areas, like many in the natural sciences where there is a need for research infrastructures may receive higher levels of project funding and this stream of resources may acquire higher strategic importance.
References


