INTRODUCTION

The French evaluation scene is founded upon what Chabbal called the guarantor model (Chabbal 1987). Even though its institutionalization at policy level dates only from the 1980s, evaluation has been a long standing issue the location and action of which changed at the same time as the research system and policy evolved. This is why, in this article, we shall start by describing the policy setting, putting the French system in perspective. Our hypothesis is that, evaluation being part of a larger managerial and decision-making frame, it is difficult to grasp the changing role and embedment of evaluation activities without referring to the transformations of the global scene. Such an approach bears a second advantage, that of highlighting the progressive enlargement of what falls under recognized evaluation activities, thus putting in perspective the too often mentioned polisemic dimension of the word evaluation. This situation, we shall see, is not linked to the fact that in the French language words like appraisal, assessment and evaluation are all translated as one generic term: ‘évaluation’.

PART 1 - THE FRENCH RESEARCH SYSTEM IN PERSPECTIVE

It is not our intention to provide a full description of the French system1 but to highlight structural dimensions which account for the specific dynamic of the French evaluation scene. The present research system can be seen as a three storey building constructed over a fifty year period. Inspired by the symmetry of French gardens, these three floors are each organised in two flats (respectively for public and industrial research) around the central staircase provided by the global research policy. But we shall see that the third floor is blurring this elegant disposition both by pushing walls down between flats and by multiplying possible staircases to

---

1See Mustar 1994
attain them.

1.1 - Two pillars: research institutions and ‘grands programmes’

The ground layer corresponds to the strong institutional choices made soon after the end of the second world war, to remedy what was perceived as a double ‘gap’.

The first gap dealt with ‘independance’ and ‘grandeur’: with the creation of the Commissariat à l’energie atomique (CEA) as early as 1945, with the aim of catching up on nuclear energy (and weapons), a lasting policy instrument was stabilized, that of the ‘grand programme’. These were later to be called “programmes de développement technologique” (PDTs). With the return of De Gaulle to power in 1958 and the creation of a high level coordinating structure to organize the national research policy (DGRST), this model was enlarged to space activities with CNES, telecommunications with CNET, industrial exploitation of deep seas with CNEXO (now IFREMER) and the lasting engagements in civil aeronautics (Concorde before Airbus) and in successive ‘plan calcul’ for the computer industry and later on information technologies. Just for the reader to bear in mind the importance of this layer and its lasting strength in French research policy, it is enough to underline that its average budgetary weight in public civil research budget still represented over 40% throughout the eighties.

The second perceived gap was scientific and resulted in the transformation of an institution limited to technical research into the ‘centre national de la recherche scientifique’ in charge of all ‘fundamental’ research in France and in charge of directly recruiting and employing full time researchers. Today with over 25000 persons including near to 15000 researchers and engineers, it represents a fifth of French public civil research capability. This approach was progressively enlarged to ‘mission oriented’ institutions from health (INSERM, over 2000 researchers) to agriculture (INRA, over 8000 persons including near to 2000 researchers) through research for developing countries (ORSTOM and CIRAD, 1500 researchers) or infrastructure (LCPC, CSTB and INRETS, some 1500 persons). This explains the first and may be foremost ‘evaluation’ activity within the French research system, this linked to career development.

Both movements extend over two decades, the last institutional creations dating back from the beginning of the 1970s. Since then, these two pillars of the French research system have channelled the core of French public support to research. When looking backwards, continuity appears as their main striking feature. There has been nearly no institutional reshuffling, the most important change being limited to the merger of the fishing research institute with the deep sea research organisation to form IFREMER! And the same nearly applies for PDTs once the failure of the

---

3even when taking university teaching staff on a 50% basis. Source : MESR 1994 and OST 1993.
computer industry actions is accounted for, with their relay taken by the European ESPRIT programmes. This may well account for the early focus of government evaluations, centered on the performance of these pillars. But this stability had another strong consequence: any significant change in research policy was either to take place within this frame or to add to it other policy instruments. This topping-up movement characterized the transformations observed in French research policy during the last two decades (from 1975 to 1995). Two major changes took place: one at the end of the 1970s pushed forward by the economic crisis and the search for greater efforts to promote innovation, the second at the turn of the 1990s with the rediscovery of the central role of universities and the fostering of the internal coherence of the research system.

1.2 - Promote innovation: technological programmes and ‘procedures’

The rising ideology of innovation as a key answer to the ‘oil shock’ and the resulting economic crisis did not only push OECD to change of terminology -moving from ‘S&T policies’ to ‘innovation policies’. It also developed all over Europe a reappraisal of public intervention. From Alvey to Esprit, a new policy instrument rose to prominence: the technological programme (Callon 1989). In France, the 1982 law on research turned this into ‘the’ main policy instrument, differentiating between ‘programmes mobilisateurs’ and ‘programmes prioritaires’. Altogether, near to twenty were launched even though nearly none mobilized anything near to what was planned in their original budgets. From information technologies to biotechnologies through new materials, they covered the usual fashionable range of domains, were managed by government departments, played a major role in the institutionalization of evaluation and, ironically, were drastically reduced at the turn of the 1990s before any evaluation could assess their effects. Only two programmes remained significant: ‘energy mastering’, probably because it was the only one to succeed in building its own institutional frame, a public agency; and ‘land transport’ linked to the strong ministry of equipment and transport and backed both by the car industry and the powerful French public transport operators. These were also the two programmes which served as pioneers in demonstrating the ‘robustness’ of evaluation exercises4.

These ‘top-down’ actions focused on strategic technologies aimed at preparing the next generation technologies; they did not answer the identified pressing need for innovation. Why were French firms less numerous in undertaking research efforts5? Why was the export pattern so

---

4See Evaluation of the PRDTTT by Centrale Management (economic effects) and CSI (strategic management) and Evaluation of AFME by CSI. Robustness refers here to the capability to resist critics from the evaluatees. Methodological approaches developed can be found in Callon, Larédo & Mustar (1995).

5Under the Frascati approach their number was estimated in 1982 at less than 1500 compared to near to 10000 in Germany. They were in 1992, 3700 (MESR, 1995) while
distorted? Whatever the reasons, debates revolved around two arguments linked to finance. Since banks were judged reluctant to share innovation risks, a new procedure had to be invented: the ‘aide à l’innovation’, managed since 1979 by ANVAR, another specific agency. This agency co-funds SME innovation projects and gets payed back when projects meet with success\( ^6 \). Since research efforts were judged too limited, a fiscal incentive -‘credit d’impôt recherche’- was developed in 1982 to grant a tax credit to all firms which increase their R&D effort\( ^7 \). These two ‘procedures’, focused on the innovation efforts of individual companies\( ^8 \), remain in operation, in contrast with the technological programmes, and today represent the bulk of public support for industrial research and more specifically for SMEs.

\[ \text{1.3- The nineties, a new focus on universities and the fostering of linkages} \]

Why did national technological programmes stop? A simple answer lies in budgetary constraints: tied with the recurrent cost of its two pillars and the political support for SMEs, programmes were the only possible source open to the government for budget cuts (all the more since their main clientele was made up of large firms and public research institutions which had already learnt to work together\( ^9 \)). Another important aspect may be found in the increasing role of EU programmes, both in term of area coverage and of actor mobilization\( ^{10} \). Support for ‘strategic technologies’, large firms and high-tech SMEs was de facto transfered to EU level. Research policy implementation was de facto delegated to the research institutions and agencies which managed both pillars through pluri-annual ‘contrats d’établissement’ between the latter and government.

\( ^{10} \)as was demonstrated as early as 1990 by the survey of the impact of EC programmes on the French research system (Laredo & Callon, 1990), a movement which has strongly deepened and broadened since (Larédo 1995).
Meanwhile, the widely recognized failure of the linear model of innovation had put to the forefront the need for more inter-action between the different parts of the system, a situation which was politically translated into two major priorities: more trained people and more connections. They organize the new storey, what the 1990s add to the French ‘national system of innovation’.

What characterizes this addition, is less the rearrangement of government departments which has taken place11 than the changing political scene. As technological programmes have been delegated to the European Commission, a similar delegation to regional authorities took place through ‘contrats de plan État-régions’. Regions, which are a recent creation from the 1960s, were the core beneficiaries of the 1982 ‘loi de décentralisation’. They have become powerful policy makers both for University development and for the support of all intermediary structures dedicated to the fostering of links with typical manufacturing SMEs, which form the core of the economic landscape and are not ‘high-tech’ enough to take part in European technological programmes12.

1.4- The French System today

Implementation of French public action in research and innovation today can thus be described as a huge delegation process: delegation to research institutions, delegation to agencies and organisations in charge of the ‘grands programmes’, delegation to ‘procedures’ like crédit d’impôt recherche or aide à l’innovation, delegation to regions through contrats de plan État-regions, delegation to the few remaining national programmes and, if one may say so, delegation to European programmes. To follow-up and monitor policy implementation, the relevance and performance of which has been in France as in most other western countries a growing concern, it is necessary to analyze the outputs and effects of all these actors which mediate between the policy making sphere and the research actors themselves. R. Chabbal has proposed to call these actors ‘research operators’ the number of which has been estimated at 200 for France alone.

PART 2 - THE PROGRESSIVE BUILDING OF EVALUATION AS A POLICY INSTRUMENT

In France, the institutionalization of policy evaluation only dates back from the 1980s and accompanies the above-mentioned delegation

---

11The ministry of research and technology, created at the beginning of the eighties to implement the innovation priority, was first joined to higher education (1993) and then included in a large department covering education, research and training (1995).

12There are more than 150 technology resource centres (CRITT in French) at present under evaluation by CNER. This added to the fact that ANVAR action takes place through regional delegations where nearly all decisions are taken, explains the growing importance of regional decision making in the framing of French public action towards innovation.
movement. It is part of a global framework for policy evaluation. Both developments took place simultaneously, but connections between the two schemes have remained marginal while their organisational setting is very different\textsuperscript{13}. This institutionalization cannot be understood without referring to the long standing organizational choices made and to the progressive enlargement of activites qualified as evaluations. On the French research scene, evaluation was first linked to the ‘professionalization of science’: how to recruit researchers and organize their careers. The ‘Comité national’ in CNRS operates a durable model for the evaluation of the ‘quality’ of research at individual and laboratory level. At the other end of the spectrum, all major changes in policy orientation have followed extensive consultations while the organized consultation mechanism of the scientific community, the ‘Conseil Supérieur de la Recherche et de la Technologie’, was required by law to provide the Parliament with an annual ‘evaluation’ of the state of the French research system. This reminds us that evaluations answer decision-making needs and that their process and scope depend on the organizational frame in which such policy decisions are prepared and taken. It has thus been felt useful for the reader to get a broader view of all these established consultation mechanisms which, as evaluations do, provide political decision makers with ‘authorized’ views.

2.1 - Evaluation, an instrument linked to the professionalization of Science

With the creation of CNRS, a new full-time profession was to develop and be managed: how should it select new researchers? How should it insure them a rewarding career? The Mertonian model of the republic of science was effective in providing operational solutions. Scientists were to be evaluated by their peers within disciplinary committees. The ‘Comité National’ was to embed this democratic approach by dividing itself into disciplinary sections, half of the members of which were nominated by higher management and half elected by disciplinary peers. For over twenty years this system followed the dynamics of science by redefining sections and increasing their number. Problems started when the question arose of the number of sections: was it possible to multiply them indefinitely? The answer was no and their number stabilized at 50. Soon a new question emerged: how to deal with researchers at the border of two of these sections? Modifying the borders did not bring any lasting answer, other researchers finding themselves between established communities. Inter-disciplinarity was felt such an important problem that all members of the Comité National and of its sections (over one thousand persons) were gathered for a

\textsuperscript{13}Following the rapport Viveret, two interministerial bodies were created for evaluating policy implementation. An interministerial committee for evaluation (CIME) was to decide upon evaluations to undertake and a scientific council for evaluation (CSE) to verify the relevance of the methodology and audit the quality of the reports. The latter produces an annual report which focuses upon ethical and methodological dimensions of evaluation while individual evaluations undertaken have mostly been kept within administrative circles. See Viveret 1989 and CSE 1995.
two-day seminar for the first time in over two decades (1990).

Until the 1980s, this system was limited to CNRS and INSERM, the two main institutions to be classified as funding ‘fundamental’ research. The 1980s, with the socialist government and the 1982 law on research, transformed all research institutions into ‘établissements publics à caractère scientifique et technique’ (EPST), established their researchers as civil servants and thus required them to be recruited by ‘concours’. The CNRS model was thus extended to all research institutions with all the hurdles linked to such an administrative transformation for thousands of active researchers. When, taking into account the reinforced role of research institutions, we must not be astonished if, in France, evaluation, for most research actors, first and foremost, brings to mind individual career issues.

2.2 - Consultation and advice at the core of policy formulation

Before research was turned into an established policy (1958) when De Gaulle returned to power, the ‘colloque de Caen’ (1956) had been effective in proposing the full framework which was to be applied later: a coordinating central structure directly under the orders of the prime minister, a central fund to support promising areas, their transformation into lasting specialized institutions once the proof of their interest was to be made, and a ‘wise-men’ committee to assist in ‘evaluating’ the relevance of the whole system and the new areas to support.

Similarly the ‘assises nationales de la recherche et de la technologie’ in 1981 paved the way for the 1982 law on research and technological development: apart from creating ‘programmes mobilisateurs’ and transforming research institutions into EPST, the law reframed the research advisory body. The ‘Conseil Supérieur de la Recherche et de la Technologie’ (CSRT) since then has consisted of some forty persons elected for a four year period through a complex representative process which led analysts to consider it as ‘the parliament of research’. It is required by law to issue every year an ‘evaluation’ of the state of the French research system for the Parliament.

But this was not enough for the Parliament. Judging that it needed its own independent source of advice, a parliamentary office for the evaluation of scientific and technical choices was created in 1984. Numerous studies have been undertaken by this office tackling all key subjects from nuclear safety to space policy, and its growing influence has been linked with its

14Civil service in France is a life employment and you can only get access to it through national “concours”. This does not only cover higher civil servants and the corresponding “grandes écoles” they come from. However you can be employed by the state (or one of its établissements publics administratifs, EPA) on a contractual basis and with a de facto life employment but without being a civil servant. Such was the case of all full-time researchers employed by CNRS and the like before the eighties.

15It was then called “délegation générale à la recherche scientifique et technique” (DGRST).
active role in the new laws dealing with medical and genetic ethical problems.

We face here a second and new understanding of the role and practice of evaluation\textsuperscript{16}. It no longer an internal mechanism to deal with the career of individual researchers. On the contrary, it only tackles global issues and gets its authority from the ‘representation’ mechanisms which establish the consultative bodies. One should not overlook this political definition of what evaluation is about and how it works, since it represents by far the most common setting for policy advice in Europe (Callon, 1995).

2.3- Between actors and system: operators central to the institutionalization of evaluation in France

Still, both government and parliament were faced with the accountability of public action. A grand programme bears visible fruits: even if they questioned their costs and/or orientations, politicians and the public at large could see Concorde fly, they could monitor the successes of Ariane or could follow the development of nuclear electricity. But when you aim at increasing the technological competitiveness of an industry, how do you measure results. At the end of the programme, you do not expect new products to be already on the market, with all the tangible assets connected to them (increased market shares, profit and employment, whatever the order of preference is). And we might guess that, if everything had been going well in the first place, nobody would have thought of undertaking a specific public programme. So how can one appreciate the transformation of the economic situation of the targeted industry? An improved situation at the end of the programme is likely to have limited sources in the programme activities while on the contrary a programme could well have been technically successful while major actors were taken over by foreign firms (as with the UK’s Alvey programme and IT industry - Guy & al, 1991).

In France this question was addressed as soon as programmes mobilisateurs were created. The 1982 law mentioned they should be evaluated while the next three year programming law (1985) had one full title devoted to the evaluation of programmes and research institutions. But no implementation structures were provided for. Our explanation lies in the negative ‘auditing’ experience of the late 1970s described below. We shall argue that institutionalization was in consequence linked to the university policy: by adopting the model of the ‘independant administrative

\textsuperscript{16}Similarly, the return of the ‘right’ to power in 1993 drove to a re-appraisal of research policy and to a similar exercise named ‘consultation nationale sur les grands objectifs de la recherche française’. The process has been the following: an 18 members committee was named and issued a preparatory report sent to all research actors and institutions (over 60 000 copies were distributed). This, together with all voluntary written submissions, formed the basis of six thematic colloquia (over 2000 persons) and a national synthesis meeting, all of which have produced reports before the final government position was issued : ‘Rapport sur la Recherche Française’ by F. Fillon, Minister of Higher Education and Research (Paris, 1994).
authority’, evaluation of universities established the frame of the French ‘guarantor’ model which was later to be adopted for all other French research operators.

**The 1970’ and the ‘punitive’ audits of research institutions**

With the oil crisis and the so-called innovation gap, questions developed on the utility and performance of applied research institutions - from transport to agriculture. A fashionable wording later encapsulated these rising doubts - non applicable applied research, ‘recherche appliquée non applicable’ or RANA in French. A wave of ‘audits’ was initiated by the French government\(^{17}\). All were on the same model. They consisted of a one man exercise without any specific support, reports were kept confidential and, on the whole, no effect on the audited institution could be observed except in the few cases where the auditor subsequently became the executive chairman of the audited organisation\(^{18}\). These exercises had a strong negative effect on both sides of the research managing community, within central government administration and within the management of the audited organisations\(^{19}\). This may well account for the reluctant attitude towards the systematisation of evaluation as a strategic management tool which was evident in the following decade.

**Successful ‘Comité national d’évaluation des Universités’...**

Meanwhile, universities were living through another round of reform as has been the case every five years since the Faure law about their autonomy following the 1968 events. The 1984 Savary law changed their status and their organisational structure was transformed. To satisfy one of the debated points about university performance, a ‘Comité national d’évaluation’ (CNE) was created. Following the success of the ‘Commission nationale de l’informatique et des libertés’, it was established as a committee independent from the ministry, reporting directly to the ‘Président de la République’. Its autonomy was ensured by a strict nomination process involving both chambers of Parliament and independent state auditing

\(^{17}\)In parallel to this, the ‘Centre de Prospective et d’Evaluation’ (CPE), managed by T. Gaudin, was created to foster both practices at government level. One of its first activities in evaluation was to undertake an evaluation of the PDTs (1981-82). It adopted the same approach as the other audits, being a one man exercise with a confidential report. This may well account for the fact that CPE stopped undertaking evaluations directly, focusing on research and information on evaluation practices. In particular, CPE was instrumental in publishing, with the support of the European Commission, the ‘R&D evaluation newsletter’ (1987-1993).

\(^{18}\)Even though an evaluation framework was being discussed, this approach was not left aside and the 1986 Chirac government reproduced a similar scheme, requesting an audit of CEA from a three person group and later nominating two of the three auditors to the two top executive positions of CEA. The new managers then circulated a public version of their report...

\(^{19}\)See the CPE seminar on evaluation practices by A. Rouban, CPE reports n° 29 and 51, Paris 1983 and 1984.
institutions, and later on, by the granting of its own logistical means and of a specific budget for undertaking evaluations. It was to evaluate all universities and was responsible for its evaluation programme, for the methodologies to adopt and for the dissemination of its evaluations.

Its success can be described in a few words. In the first four years of its existence, it evaluated some 40 universities, all of which had volunteered, underlining the quest for external and periodic scrutiny and advice.

To manage such a programme it rapidly established a clear methodology. Each evaluation is under the responsibility of two members of the committee who will write the evaluation report. Universities are required to prepare in advance a preliminary ‘information package’ and a set of ‘indicators’\textsuperscript{20}. Numerous disciplinary experts (between 10 and 15 per evaluation) are called upon to evaluate the different departments and, in order to ensure their freedom of judgement, the CNE has chosen to keep these experts’ reports confidential. The final evaluation report has to be agreed upon by the full CNE, it is then sent to the University which provides for a written answer. The report and the answer together constitute the evaluation file which is published in what is now known as the blue collection, because of the color of their covers. The annual reports to the Président de la République are used by CNE to deal with transversal issues and thus address policy formulation.

... Paved the way for the periodic evaluation of other research operators

As mentioned earlier, the 1985 law which made evaluation a compulsory feature of the RTD policy did not define any implementation structure. To overcome the above mentioned attitude, experiments were undertaken with voluntary operators (some five programmes and one agency)\textsuperscript{21}. Coupled with the success of CNE, this paved the way for defining an operating frame for evaluation. At first, it was contemplated to enlarge the responsibilities of CNE, but after four years, despite its success, it had only evaluated a third of the French Universities (it has taken CNE more than a decade to do the full round and start a second one). Government thus opted for replication. The Comité National d’Évaluation de la Recherche was created in 1989 on the same basis as CNE: independent from the ministry of Research, made up of 10 members nominated for 6 years after a strict process involving Parliament, auditing bodies but also the French Academy of Science and CSRT, annual reporting to the Président de la République, its own logistical means and own budget, complete freedom in the choice of its calendar, its methodologies and its dissemination policies. One of the key questions was the definition of the operators which CNER was to systematically and periodically evaluate. For once a pragmatic solution was adopted by including in its scope all institutions and R&D

\textsuperscript{20} As expected this has been the last part of the evaluation process to be stabilized, the definition of these requirements and the corresponding indicators being still the object of methodological research at the beginning of the nineties.

\textsuperscript{21} see Mustar & Larédo in Callon (1995) and Callon & al. (1992).

10
activities which received state support from civilian budget\textsuperscript{22}.

\textbf{PART 3 - CNER’s approach to evaluation: a four step process}

CNER started operation in 1990 and now has five years of experience. Since it is a unique feature in the European evaluation landscape, the authors have chosen to devote the last section of this paper to a description of the ‘guarantor’ model at work, to an analysis of its effects and to drawing preliminary lessons. Describing the agenda setting process and the four-step evaluation process adopted will help in capturing both CNER’s main effects and its limitations.

3.1- The agenda setting

With sole responsibility for its agenda, and having to evaluate all operators in a six year period, CNER was faced with an initial problem: in what order to proceed? and in particular what relations to establish with the immediate political concerns? The decision was made to answer to all requirements from ministers (in practice only two cases in five years) and to have an equilibrium between large and small spenders, and between the main types of operators - research institutions, programmes and agencies, procedures. For each of them, choices were made on specific grounds, highlighting the diversity of ‘research operators’. New experimental procedures (such as CRITT) were preferred to more established ones, such as the research tax credit or the ANVAR innovation support to SMEs. The choice of programmes and agencies took into consideration the existence of internal evaluation efforts or recent ad-hoc evaluations; it focused on ‘programmes mobilisateurs’ (biotechnology and working conditions) and selected different implementation configurations - an agency (space ‘grand programme’), an inter-ministerial arrangement (new materials). And the core of the efforts was dedicated to French strongest pillar, research institutions, CNRS being dealt with not globally but by research area.

After five years of activity, CNER is far from having covered all French operators. During the 1994 ‘consultation nationale’ (see part 2.2), questions were thus posed about the adequacy of the periodicity of such external examinations. One of the six main topics around which this consultation focused, dealt with French research structures and included a specific workshop on evaluation, highlighting the role evaluation has now taken in the organizational frame of the French research system. Most of the issues discussed referred to the evaluation process adopted by CNER and its effects.

\textsuperscript{22}State support being defined by the 22 ministries which signed the “decret”. The road was left open for other institutions - private or military - to ask CNER to evaluate them or for ministers to ask CNER to consider the evaluation of any organisation funded by their department. See CNER 1994.
Its four steps will be presented now before returning to global effects and limitations.

3.2- Define the terms of reference: a crucial and key dimension of the guarantor model

How to proceed, was the second question CNER had to cope with. A standard process was looked for, but this did not mean that evaluations should face similar questions. Should then terms be determined at the outset of the evaluation, based on the objectives or missions of the evaluated operator? Of course, one would be tempted to answer, but this was not felt very pragmatic for institutions or programmes which existed for decades. Furthermore, CNER made it clear in its first annual report that it wished to adopt a ‘pro-active’ approach, not focusing on initial objectives and subsequent potential divergences, but on present problems and issues through which to re-read recent action and its effects. The definition of adequate terms of reference (‘cahier des charges de l’évaluation’ in French) was thus an integral first part of the evaluation process.

To be in a position to establish them, information had to be gathered on the evaluatee, the main stakeholders identified and discussions organized around key issues and problems so that terms of reference could be directed to them and methodologies duly selected. CNER called this the ‘characterization’ phase.

This phase rapidly turned into a very time-consuming task which required specialists, ‘evaluation professionals’. The global information production process was managed by internal professionals, CNER’s ‘chargés de mission’, while, increasingly, external professionals were called upon for specific aspects (mailed surveys of actors mobilized, economic analysis of direct and indirect effects, institutional analyses, financial and budgetary management audits, etc). This work, and the discussions both with the evaluatee and the main stakeholders (government, sectoral partners, etc), was undertaken under the supervision of two or three members of the Committee in charge of the whole evaluation process.

The formalization of this ‘preliminary’ stage, before evaluation questions were formulated, is probably the key dimension of this ‘guarantor model’, a dimension which has had the most lasting effects. It first highlighted the lack of managing and monitoring instruments and was, in many cases, instrumental in developing tools which have later been

---


internalized by the evaluees: evaluation could thus be mobilized for reviewing the relevance of managerial practices. It had a second, even more powerful effect, in changing attitudes towards evaluation. It was no longer perceived as a sanctioning mechanism coming from ‘out there’, but rather as a negotiated reflexive exercise on issues which were identified as crucial for the coming years. It thus appeared less an ex-post exercise on past objectives than a strategic assessment of current problems and issues.

3.3-Robustness at stake: the expertise and confrontation phases

What is often referred to as evaluation - the work done to prepare an answer to the terms of reference - was dealt with entirely externally, mainly by individual (thematic or disciplinary) experts (worldwide) and some reliance on well established methods (mailed surveys...). CNER has thus not been innovative in evaluation methods. However the choices made for this ‘expertise phase’ do not correspond to the usual practice of evaluation panels (Bobe and Viala, 1995). Experts are connected to individual problems specified in the terms of reference and are required to produce individual ‘topic reports’, the latter considered as complementary building blocks in the evaluation ‘dossier de base’. In order to avoid a collection of individual answers and to promote a coherent global frame for action, experts are also asked to produce a collective report as their main output (‘rapport d’instruction’ in French) which serves as a basis for discussion with the evaluee. CNER has given prominent importance to this ‘confrontation phase’. The report is circulated to the evaluee and written comments gathered. This provides the basic material for a dialogue between experts and representatives of the evaluated body25 in front of all ten committee members.

This dual approach to the expert report was effective both in ensuring, through individual reports, an effective treatment of all questions and in organizing, through the collective report and the necessary compromises experts had to enter with one another, a coherent approach to the evaluated operator as a whole. Even though such a process faces similar problems about the choice of experts as do all panel based evaluations (Bobe and Viala 1995 and Massimo 1995), it must be underlined that up to now, the information base has never been challenged and the experts’ analyses and proposals never rejected as irrelevant. Some aspects of the latter (especially those relating to the global framework proposed by experts) could be challenged during the confrontation phase which has been, to use CNER’s expression, ‘very fruitful’ and instrumental in promoting ‘acceptable’ or ‘realistic’ recommendations. This dual step process can thus be considered as effective in building robust evaluations.

3.4- Direct effects at operator level and indirect effects at policy and system

25When it is a research institution, both management and workers representatives are confronted with experts.
The evaluation process does not stop with the confrontation. The Committee decided that it should capitalize on this confrontation and make its own recommendations in a short synthetic document written, and unanimously adopted, by the Committee. This constitutes the only official document issued by CNER\textsuperscript{26}, its ‘recommendations’ which are public and are, as its annual report, regularly printed by La Documentation française, the French H.M.S.O.

This process and especially this last phase requires a heavy involvement from members\textsuperscript{27} and explain to a great extent the observed lower than expected speed (which was to be ten evaluations per year when fully established). In five years of activity CNER has undertaken 14 evaluations and finished 11, which amounts to an average of three evaluations per year\textsuperscript{28}. Even when taking into account the start-up period, we are faced with five evaluations per year and, like CNE, with over a decade before the same operator is evaluated again. Coverage appears under the present working frame a first limitation.

The analysis of the effects of evaluations brings to the fore a second limitation. Evaluations have been quite influential at operator level, especially with the renewal of management tools. However, CNER’s recommendations have had limited influence on policy making and, when proposed, on rearrangements of the institutional frame. This conclusion which was put forward during the 1994 ‘consultation nationale’ and applies for both committees, highlighted the absence of any take-up mechanism. How to ensure that recommendations are dealt with and reasons made

\textsuperscript{26}All other documents are not endorsed by CNER and remain as grey litterature, even though some have been heavily circulated.

\textsuperscript{27}Members meet half a day once a week. With the follow-up of individual evaluations and the writing of recommandations, it was calculated that this was an effective quart time job, making it difficult under the present conditions to increase the number of evaluations unless through changing either this last phase or the rules on which the Committee is established - Membership in some “administrative independant authorities”, such as the national committee for the independence of radio and television, is considered as a full time activity.

\textsuperscript{28}These 14 evaluations are as follows. Half deal with research institutions: CSTB on construction, IFREMER the sea research institution, INSU and SPI two of the seven scientific areas of CNRS, ORSTOM and CIRAD, the two institutions dedicated to (under)development. Four deal with programmes: space, biotechnology, new materials and ‘technology, employment and working conditions’ (out of which three are now extinct). Two procedures only have been evaluated: CRITT, the French Technology resource centres which have been developed under the umbrella of the ‘contrats de plan Etat-regions’, and GIP (groupements d’intérêt public), a specific setting created by law to organize on a special problem and for a limited period of time, numerous private and public efforts, and a setting which is more and more used for organizing problem solving actions or programmes such as those on AIDS or on human genome. Finally the last on-going evaluation is undertaken jointly by CNE and CNER and focuses on the interactions between universities and research institutions in life sciences. Recommendations issued up to 1994 are gathered in CNER, 1994b.
explicit when they are not followed? Is the follow-up process initiated by CNER itself\textsuperscript{29} adapted? One proposal was to link this evaluation process with the existing consultation mechanisms (see part 2.2) and to make it compulsory for the evaluatee and the government to produce a follow-up account to the ‘Office parlementaire des choix scientifiques et techniques’, thus bringing to the legislative level the debate about potential rearrangements within the national research system.

By disconnecting evaluation from immediate policy and government management concerns, a gap is created which has two contradictory effects. One is to focus the evaluation on effective long term issues and thus promote the ‘pro-active’ and ‘strategic’ dimension of the evaluation exercise. This may be highly relevant for evaluated operators as long as they can take it on board themselves. But, whenever, policy issues were at stake, it has proven difficult for them to enter the policy making process. This demonstrates that the guarantor model is effective in producing robust and strategic evaluations. But it also underlines the need for an institutional frame to cope with their insertion in the political decision-making process. This does not mean to ensure that their conclusions and recommendations are enforced but that they are discussed at the relevant policy level. This appears as a key issue for the full articulation and the future of the French evaluation scene; all the more since this institutionalization has had strong indirect effects on the internal practices of many French ‘operators’.

3.5- Institutionalization as an incitement to ‘strategic’ management

Just as it is not enough to grasp only the direct effects to analyse the impact of a programme, so for an evaluation, indirect or structural effects should also be accounted for (Bach 1995, Laréd\textsuperscript{o} 1995). Though causality may be questioned here, we see three main changes in the French system over the last few years which have been, at minimum, helped by both committees’ actions and repeated recommendations.

The first one deals with the lack of management tools. The committees’ procedures had a strong effect in pushing research institutions and operators to prepare themselves for evaluation and thus design internal processes which have since had, at least in some cases, a clear impact on managerial practices. The ‘Delegation des audits’ created by CNRS to report on issues crucial to top management has been very active and, through a very rigorous process controlled by its own guarantor, the ‘Comité des audits’, has produced evaluations dealing as much with logistical issues (the CNRS computer frame...) as with interdisciplinary research programmes or, more recently, with the functioning of the ‘Comité National’, this crucial body managing the researchers’ career (see part 2.1). Evaluation reports are public and include the answers of CNRS management and the corresponding transformations which are planned. Some applied research

\textsuperscript{29}CNER has by law the possibility of monitoring the results of its reports. It aims to produce by the end of 1995 a special report to the ‘Président de la République’ on this subject.
institutions like CEMAGREF, or research agencies like ADEME, have developed functional ‘evaluation and strategy’ directorates and government departments, such as the one in charge of equipment, transport, construction and housing research, have developed their own evaluation unit (which, in the above-mentioned case, recently finished a new evaluation of the Land transport programme). Slowly evaluation is embedding itself in what is the core of the French research system, its lasting research institutions and agencies.

The second indirect effect is linked to the pro-active approach adopted. This was another way to emphasize the need for a more long term, ‘strategic’, analysis to complement the over-dominating role of annual budgeting. This was the ambition of the four year ‘contrat d’établissement’ which the ministry of education started with universities in 1989, after five years of action and evaluations by CNE. These pluri-annual ‘contrats d’établissement’ have also been established as a major implementation mechanism for French research policy. And they join the five-year ‘contrats de plan Etat-régions’, the growing importance of which has been underlined in the first part of this paper.

A third indirect effect was to underline aspects which required further methodological research like industry-public research links or renewed analysis of laboratory dynamics. But CNER has not been actively involved in any methodological research and most of the work done by French teams has been supported by some French operators but mainly derives from their participation to European and regional evaluations.

CONCLUSION

We have seen that the embedment of Evaluation in the French research system was first linked to the professionalization of Science and the corresponding need for a management process for both recruitment and career development of full-time public researchers. We have also seen that ‘Evaluation’ was a term long connected with all the advisory mechanisms which have accompanied the framing and evolution of French research policy. And finally we have shown that the institutionalization of evaluation, i.e. the creation of specific public structures devoted to this activity within the research system, was closely linked to the delegation process adopted for policy implementation and the major role of powerful ‘operators’. It is only in this last and recent frame that evaluation practice developed in France as an articulated process based on proven

30Among others, works by BETA on indirect effects of programmes (first supported by ESA and enlarged through the evaluation of Brite programme), by CSI on techno-economic networks and structural effects (supported by ADEME and then by the EU - evaluations of NNE3, JOULE and MHR programmes, impact studies) or on laboratory profiles supported by CEA and by regional authorities... Most methodologies recently developed in France have been presented in Callon & al 1995.
methodologies and corresponding to what science policy researchers and evaluation professionals tend to define as ‘effective’ evaluation practice.

Our first conclusion bears thus upon this definition. We should not restrict our approach to evaluation to an articulated set of practices dedicated to the ‘on going’ or ‘ex-post’ dissection of managerial practices. This focus on implementation derives from the need for accountability which has been, in the 1980s, the booster of evaluation practices at government level. But when taking their main output, recommendations and advice for decision making, we have shown in France the role of other settings which, because of similar outputs, were considered as evaluation ‘providers’. In a way, this is what recent interventions by researchers from the Fraunhofer Institute for innovation (Becher & Kuhlman 1995, Meyer-Krahmer, 1992) are promoting when speaking of the need to better articulate ‘ex-ante’ evaluation exercises. Furthermore, the French scene highlights the fact that these proposing bodies embrace more than implementation aspects and tend to focus on global issues, so that, at least in France, they appear as a key circulating instrument for the take-up of evaluation findings and recommendations.

It might be that this issue is linked to the specificity of the French evaluation scene which illustrates the guarantor model at work. The detailed review made of its functioning, bears, from our point of view, three main lessons. The guarantor model is a very powerful mechanism for giving evaluations their full role in decision making by turning them towards being ‘pro-active’ and making them become key strategic exercises. It is also instrumental in delivering robust evaluations which stand criticism and can serve as a basis for a renewed strategic and political debate. But for the latter to take place, it requires careful thought with regards to the circulation of evaluation results and recommendations. Publications are not enough, nor is the strong influence that such evaluations have had on the managerial board of evaluated operators. A policy take-up mechanism has to be devised to embed this very costly and time consuming advisory exercise into the policy decision making process. This is still not the case in France where connection is being sought with other advisory committees which both the executive and legislative bodies have developed to help in framing research policy.

Bibliography


Ministère de L’Enseignement Supérieur et de la Recherche, *Recherche*


**Glossary of French research organisations and committees**

ADEME Agence de l’environnement et de la maîtrise de l’énergie
ANVAR Agence nationale de valorisation de la recherche
CEA Commissariat à l’énergie atomique
CEMAGREF Centre national du machinisme agricole, du génie rural, des eaux et des forêts
CIRAD Centre de coopération internationale en recherche agronomique pour le développement
CNE Comité national d’évaluation (des établissements d’enseignement supérieur)
CNRS Comité national d’évaluation de la recherche
CNES Centre national d’études spatiales
CNEXO centre national d’exploitation des océans
CPE Centre de Prospective et d’Evaluation
CRITT Centre de recherche, d’innovation et de transfert technologique
CSE Conseil scientifique de l’Evaluation
CSRT Conseil supérieur de la recherche et de la technologie
CSTB Centre scientifique et technique du bâtiment
DGRST délégation générale à la recherche scientifique et technique
IFREMER institut français pour l’exploitation de la mer
INIRA institut national de la recherche agronomique
INRETS Institut national de recherche sur les transports et leur sécurité
INSERM institut national de la santé et de la recherche médicale
LCPC Laboratoire central des ponts et chaussées
ORSTOM institut français de recherche scientifique et technique pour le développement en coopération