

## **Ethics and Evaluation**

(Version 2 - December 2004<sup>1</sup>)

Evaluation is an integral part of scientific research, especially in the public sector. Reviewing a paper in a scientific journal, recruiting a researcher, deciding whether or not to build sophisticated instruments or apparatus, inspecting laboratory activities, responding to calls for tender - all these tasks involve critical examination, appreciation of originality, estimation of costs and risks and gambles on the future, and therefore entail the comparisons and choices that are at the heart of evaluation processes.

The remarks which follow are the initial step in an ethical reflection on the function and conduct of evaluations within the CNRS. Our primary aim has been to consider how evaluations should be done and, with this in mind, to discuss the values, principles and rules that best suit the purpose. The aim is to launch a process that should eventually lead to greater fairness and efficiency in selection processes and scientific policy.

We began by looking at the principles and standards of behaviour to which evaluators should adhere, whether as individuals or groups. But ethics is not only a matter of promoting an ideal model of behaviour which those involved in the evaluation process should observe. It is also important to ensure that the positive effects of people's moral attitudes are not cancelled out or distorted by explicit or implicit rules of organisation, in other words by a system which, for structural reasons, is unable to preserve the essential objectivity and fairness of an evaluation. Even the most demanding personal standards of objectivity and impartiality are not enough when the rules of organisation allow a margin for bias or obstruct the introduction of fairer procedures. This is why we also looked at the way evaluations are organised within the CNRS. The third aspect we addressed concerns the legitimate demand made by those being evaluated for evaluations that are not only fair, impartial and frank, but also pertinent and adapted to specific situations.

### **I - Evaluation**

#### **1. The different types of evaluation**

Evaluations are of two main types:

- The first concerns the objectives and programmes proposed by researchers, research teams, research organisations or scientific bodies. They may be described as **strategic** evaluations and take place in three stages: the “*ex ante*” stage, in which an organisation or team identifies the objectives it is seeking to achieve; a second stage that analyses the way in which the objectives of a research activity are achieved (sometimes called “monitoring”), and finally, the “*ex post*” stage which reviews the

---

<sup>1</sup> CNRS Ethics Committee (COMETS) - version 2\* including a paragraph on the professional ethics of evaluation (page 3) – (December 2004)  
Version 1 was published in June 2004.

final output of the activity, in particular by comparing the results obtained with those initially set out. Strategic evaluation is associated with operational evaluation, which seeks to identify the personnel and the means required to launch (or re-launch) a research operation.

- The second type is **scientific** evaluation proper, which analyses the output of a researcher (or group of researchers) and the influence of that output within and beyond the relevant discipline. In this area, peer review has become the international norm. Although this carries the risk of encouraging pressure groups, and despite repeated attempts to find alternative methods, based on the predominance of numerical indicators for example, peer review has so far proved irreplaceable.

## **2. The ethical dimensions of evaluation**

All types of evaluation have an ethical dimension in so far as they lead to decisions that affect the lives of researchers. But they also have an impact on the relationships between the various responsible bodies within research institutions and on relationships between the research community and the rest of society.

Analysing the ethical dimensions of evaluation is problematical because evaluations rarely take place by themselves : each one is nearly always conducted within a network or a series of other evaluations. This interdependence is illustrated by three typical situations:

- Evaluations are based on the current state of knowledge. But the knowledge to which the evaluator has access is often in the form of publications. Because of this, it will already have been filtered by editorial board reviews;
- Evaluations refer to objectives which have been previously defined either by the researchers themselves or by a research body and/or external clients. These objectives could not have been fixed without a critical evaluation based on what was known, on the means of achieving the objectives in question and on conjectures that compare several distinct possibilities or take a different approach;
- Analyses of human and technical resources that can be called upon to achieve a given objective require an evaluation that is not limited to a numerical description but which, through results obtained previously, allows a projection into the future based on an estimate of individual potential.

Evaluators are therefore not alone in conducting their evaluations. They have to be aware of the weight of previous evaluations which may influence their own, which may have been debased (through no fault of their own), and which, especially, may no longer be relevant to the situation at the time of the new evaluation.

Another problem is that researchers can be either judges or judged (they are frequently evaluated, but they evaluate other researchers just as frequently) and they work within networks of relationships whose importance must not be underestimated.

Evaluations, therefore, are always complex processes. An additional difficulty lies in the fact that the most important discoveries are often “surprise discoveries” that result in sometimes agonising reappraisals of earlier certainties, upsetting ongoing research processes and even established hierarchy. To ensure that such occurrences, which are admittedly rare but

extremely important to the continued renewal of research, are actually allowed to happen is one of the aspects that need particular attention in organising an evaluation.

### **3. Deontology in evaluation**

Deontology, often called professional ethics, is a term that has several meanings, the first of which, drawing on philosophy, defines it as the science of duty - as opposed to ethics, which is concerned with what is right, or to morals, which draws boundaries between what is allowed and what is forbidden. Here, we use the most common legal acceptance of the term (i.e., “all duties that are inherent to the exercise of a professional activity”), which makes deontology subject to disciplinary regulations. There is a third meaning, sometimes described as “neo-deontology”, which has no legal value but in which charters or codes of good practice serve as references in specific contexts (like the “*netiquette*” for Internet use) and help to establish trust.

The two latter meanings, which need to be viewed in light of the specific features of scientific research, have distinct components: on the one hand, the formal professional rules stipulated by the institution which has the authority to issue and enforce regulations, and on the other hand, the codes of good practice with which a community, in a specific situation, regulates its own practice, creates trust and strengthens its credibility among its internal or external partners.

Concerning the evaluation exercise in particular, which is fundamentally based on the pre-eminent role of peer review, the community of evaluators - who, as we said, are also subject to evaluation - cannot be likened to a group of “evaluation professionals”, like an auditing agency for example, and is therefore not covered by regulations on professional status.

Evaluators will surely recognise their approach in the purely ethical principles put forward in this memo, and should see it simply as an invitation to determine their own approach. Nevertheless, it is clearly necessary to define a position which is explicit, collective and known to all concerned. It therefore seems more realistic, where individual and collective behaviour during an evaluation is concerned, to emphasise the third meaning of the term “deontology”. Evaluators and heads of establishments are encouraged to fix their own rules of conduct, by means of a charter of some kind. These however, do not imply any actionable legal obligations. The obvious aim is to strengthen confidence in the evaluation process and in the opinions delivered. In establishing these rules, the organisations concerned (National Committee and the various selection committees) cannot ignore inherently ethical criteria, such as those put forward for discussion by the COMETS.

### **II - Standards of behaviour and duties of evaluators**

Whatever its objectives, and however rigorously it enforces procedures that are known and accepted by all concerned, no institution is immune to dysfunction and abuse, and it is the role of ethics to analyse these (see previous COMETS report on “Ethics and Institutions”, 1997). Dysfunctions and abuses stem in particular from practices that distort the rules of normal and healthy competition between researchers and research teams. These practices become established all too easily, simply because objective criteria for a scientific evaluation are sometimes difficult to define and even more difficult to apply. This is why evaluations, even though the people conducting them are often unaware of the fact, increasingly tend to:

- 1) Give prominence to questions of reputation which, by definition, more often reflect past situations than situations in the present or future. In particular, the effects relating to reputation establish a form of inequality between researchers, who will not all have the same capacity for attracting attention to their work, their disciplinary field or their projects;
- 2) Call on experts who, however patent their concern for objectivity and neutrality, are either unlikely to cover the entire disciplinary field in question, or belong to networks and affinity groups that are not merely scientific;
- 3) Use a “coded language” that abounds in seemingly laudatory clichés that are actually negative and not conducive to clear criticism or constructive comment.
- 4) Take up a great deal of time even though the experts may neither wish nor be able to devote much of their time to evaluations. Evaluation means reading the entire output of one or more people and reviewing the project or team result against its global scientific context to appraise its originality or value, which, it must be admitted, can be a virtually impossible task. Moreover, it is not a particularly gratifying task, a fact that also tends to discourage some of the more highly qualified members of the disciplinary field concerned, especially when they are expected (as they should be in every case) to refuse any opportunity for taking credit from their presence on an evaluation board.

In view of these comments, we suggest that evaluation practices should be governed by a set of rules.

### **1. Confidentiality is an obligation**

The fact that evaluators have access to confidential information means that they have an obligation to be discreet and must not use the information communicated to them for evaluating researchers, articles or research projects for their own purposes or communicate it to others.

Evaluators should have no reason to fear that their statements may be communicated, in full or not, to the outside world by other members of the evaluation committee. They must not be subject to any pressure from those they are evaluating, which is the main argument in favour of ensuring that the experts on reading committees remain anonymous, although they may identify themselves if they so wish.

Guarding the anonymity of those who defend or criticise candidates, articles submitted or research projects, has a corollary, namely, that committee members must ensure that the composition of evaluation boards, editorial committees and so on is made public.

### **2. Explicitness is a necessity**

An essential requirement in any evaluation is explicit presentation of the criteria and reasons for eliminating, recruiting or promoting candidates and for rejecting articles or refusing resources. It is only on this condition that those being evaluated will understand and accept the verdict of the evaluation board. This implies that the question put to evaluators by the authority requesting the evaluation must be fully explicit.

The evaluators must justify their conclusions so that they are able to withstand scrutiny in the event of an appeal. They must give specific and well-argued reasons, avoiding any innuendo or ambiguity in their judgments.

### **3. Limiting the effects of influence groups**

Any evaluator inevitably belongs to several communities (disciplinary, geographical, ideological, etc.). They must not allow these affinities to interfere with their evaluations. The problem arises in particular with recruitment boards, which must not tolerate any form of favouritism, whether towards unions, disciplines, mandarins or geographical areas.

The same problem arises in strategic evaluations. While it would be difficult for any evaluator to recommend a reduction or freeze in funds awarded to their own discipline, they must be able, if necessary, to support emerging disciplines or disciplines other than their own.

Finally, the problem of influence can arise in reviewing scientific manuscripts, when the results submitted contradict prevailing theories or conceptions in the “school” to which the expert belongs.

### **4. Ethical responsibilities of evaluators**

The consequences of an evaluation for researchers or research teams whose submissions may not have been treated satisfactorily can be far-reaching and sometimes dramatic. This fact alone illustrates the burden of responsibility weighing on evaluators. It is a responsibility that demands perseverance in seeking information, in comparing different sources and in meticulously studying the available documentation. It requires evaluators to be prudent, free of prejudice and capable of revising their judgments in the light of new information. It also often demands courage in resisting pressure that may be exerted by those being evaluated, heads of organisations, politicians and even public opinion.

Ideally, the authorities requesting an evaluation should not merely call on each evaluator’s sense of responsibility, but should also ensure that these responsibilities are exercised under trustworthy conditions. To achieve this, mechanisms are needed to detect any abuses and, should they occur, to introduce and enforce a range of sanctions. An evaluation should never be requested from a person who has a direct interest in the elimination or selection of a candidate. For the same reason, an evaluator in the same situation should refuse to carry out the evaluation.

## **III - Ethical problems and evaluation boards**

The various boards carrying out different types of evaluation at the CNRS have evolved during the organisation’s history.

It might appear that reflections on CNRS evaluation boards are not a matter for an ethics committee but rather for policy and union officials, heads of organisations and the scientific community in general. However these reflections have an ethical dimension in so far as some organisational aspects of evaluation boards, which at first sight seem efficient, simple, rapid or “democratic”, can actually give rise to ethically equivocal situations.

Although it is not within the remit of COMETS members to make pronouncements on choices concerning thematic boundaries between CNRS sections or voting methods (lists or individual candidates) for example, or appointments committees (one or more), we believe that the authorities responsible for these choices should consider the ethical implications of each organisational method. To take this further, a number of recommendations are given below which we believe could help avoid distortions that are beyond the evaluators’ control.

## **1. Guaranteeing the scientific quality of evaluators**

The peer review process, the only universally adopted form of evaluation, implies that researchers are evaluated by representatives from their own scientific community. The question then arises as to which criteria should be used to organise researcher representation, given that the question can be formulated in different ways depending on the nature of the scientific community under consideration.

Scientific excellence should be a priority criterion in choosing evaluators, even though there are others. The objection could be made that the most able scientists are not necessarily those most capable of making fair judgments, and that excellent scientists do not necessarily have the objectivity or open-mindedness required for an evaluation. It should also be emphasised that scientific ability is an ambivalent concept that does not have the same meaning for all disciplines. The scientific community may not be in unanimous agreement as to what it understands by “ability”. Moreover, a strategic evaluation requires overall vision, a true grasp of the global picture, over and above divisions between disciplines, and this does not always go hand in hand with in-depth knowledge of a scientific field. The very principle whereby “the most able” is chosen is therefore difficult to put into practice. But however difficult to apply, this is the only universally recognised principle and must take precedence over any other. This is why the scientific ability of evaluators, whether they are elected or appointed, has to be beyond dispute.

To ensure that this is indeed the case with elected members, the composition of the electoral college needs to be given some thought. Some research organisations, like the INSERM, have restricted the electoral colleges of evaluation boards to researchers and lecturers receiving support from the organisation. The “national” status of the National Committee, and especially its responsibilities for reviewing university activities, does not allow such restrictions. However, it is worth considering whether, when electing evaluators, the National Committee’s electoral college should centre on the community of those who are actively engaged in a research activity, whether they belong to the CNRS or not, since it is crucial that academics and scientists from institutions other than the CNRS should participate in evaluation boards.

The presence of nominees is supposed to balance election results (especially any under-representation of disciplines, schools or minority trends within the French research community). The practice also brings scientists from other horizons onto evaluation boards, from industry or foreign universities for example. Although these checks and balances are legitimate, the practice of nominating members has sometimes attracted criticisms of two kinds: lack of clarity and confusion between evaluation and decision levels. Lack of transparency in the mechanisms and motivations of nominations can discredit a process whose validity in principle is not disputed by the scientific community. The point should also be made that even if a decision is based on an evaluation and even if, when analysing a scientific field, there is always a risk of confusion between advisors and actors, the nomination of evaluators should never be perceived as guided by a desire to confirm choices that have actually been made beforehand by decision-makers.

## **2. Re-examining ways of reinforcing the role of young researchers in evaluation committees**

When it was decided that colleges A and B should have an equal number of elected representatives, the measure was clearly an effective response to calls for better representation

of younger generations on evaluation boards. Nevertheless, the best researchers from B often decide not to participate in the National Committee (as elected representatives or nominees) because of a legitimate ethical concern to avoid being both judge and judged for promotions (since they are the ones with the best chances of receiving promotion), but also because of the heavy workload involved, which can be a real handicap for their research activities. We feel there is a need at this point to launch discussions on procedures that would help increase the participation of the most dynamic young researchers in evaluation boards. Such measures should, amongst other points, aim to reduce the workload of evaluators, for example by sharing National Committee functions between several committees.

### **3. Reducing conflicts of interest**

There is a need to introduce explicit procedures that are known and agreed to by all concerned to deal with conflicts of interest (for example, when an evaluation deals with the promotion of someone belonging to one's own laboratory, discipline or union).

Even though what little specific information available to us on the self-attribution of posts in the current system indicates that the practice is limited in extent, the risk of abuses of this kind, which are severely criticised in the scientific community, weakens the credibility of any evaluation board. It is therefore essential to introduce "technical" measures to avoid these problems. Calling upon people's personal sense of ethics, which is a necessary though not sufficient condition, is not enough. The rules defined should be realistic, so that their requirements do not make impossible demands on individual ethics.

The measures to be taken could be of several types. Some could involve a change in regulations, particular the rule requiring the members of a board to attend all its deliberations, including those which concern the recruitment of a researcher for their own laboratory. Others might involve a review of the composition of evaluation boards. In cases where the scientific community is large enough, one solution adopted in some countries involves two boards covering the same area of competence, with assurances that dossiers submitted by the laboratory whose members are on one committee are always assessed by the other one. Other possibilities might be to separate the boards responsible for evaluating laboratories from those handling recruitment or promotion, or reducing the length of time members have to serve on boards but forbidding them from submitting a dossier while they are doing so. Although all these proposals have advantages and disadvantages in areas other than ethics, we believe that in any discussion on reorganising evaluation boards, it is essential to take these ethical dysfunctions into consideration.

### **4. Consequences of restrictions**

Evaluations must not be made to say what they cannot say, especially when possibilities for promotion are much fewer than candidates with recognised ability. Restricting posts will never lead to a fairer selection, and the same is true for the allocation of resources. If restrictions last for a long time, it can permanently affect the credibility of evaluation boards and give rise to types of behaviour that may damage the quality and transparency of their *modus operandi*. In any situation of this kind, responsibility rests with decision-makers, not with evaluators.

#### **IV – The rights of those evaluated**

Young researchers can have legitimate career aspirations. Unfortunately, their expectations can also be unfounded or exaggerated. Researchers therefore need to be helped to review their situation and abilities lucidly and objectively. This means they need to have access to the evaluation results that concern them, in line with customary practice in public service. More generally, it would be desirable for researchers to discuss matters with their evaluators at different stages in their career. To increase researchers' acceptance of the various evaluation processes they are subject to, specifying legitimate means of appeal and mediation may be a promising idea. Procedures such as these already exist for submission of research articles to scientific journals.

Explicit provision needs to be made for cases where scientists cannot be put into a given thematic category. There are numerous historical examples of leading researchers whose contributions cross thematic boundaries quite unpredictably, and it hard to imagine that this will not continue to be the case. Researchers should therefore have the right to request individual assessments that cut across conventional sections, so that their creativity can be assessed outside the boundaries that are legitimately defined to handle the majority of dossiers.

Other researchers may lie outside the usual categories because of non-standard behaviour (biased publications, relational problems, opposition to prevailing ideas, etc.). It is often difficult to make a distinction between the consequences of an atypical psychological profile and inappropriate behaviour due to excessive originality. Past experience and morality both plead in favour of giving the benefit of the doubt to originality. However, if evaluators become quite certain that, despite the justifications given for their opinions, a researcher is taking undue liberties with the independence he is entitled to by spreading obvious errors, they must have the courage to take disciplinary measures.

#### **General conclusion**

The principle whereby any researcher (or research team) should only be evaluated by peers and with the greatest possibly transparency needs to be reasserted with force. The attention of evaluators should systematically be drawn to the risks of abuse they are exposed to and/or may succumb to. We are aware of the fact that imposing norms such as these will not immediately affect people's behaviour. However, making the norms explicit and, especially, giving clear justification for their introduction, is the first stage in the long-term process of interiorisation that will have a real effect on behaviour, at least in the sense that a person whose behaviour deviates from the norm cannot escape being aware of the fact.

Evaluation always involves a certain amount of risk. Because of the unpredictable nature of fundamental research, we would be inclined to suggest highly flexible evaluation procedures based on trust and justified by the originality of earlier research team results, at least as much as by their compliance with the institution's stated - or assumed - priorities. More generally speaking, it is essential in any evaluation process to detect and encourage the emergence of the unexpected (whether in the work of individuals or teams), to preserve every opportunity for explorations off the beaten track, reject the clique mentality and keep a watchful eye out for over-cautious or conformist attitudes.