



UNIVERSITÉ
PARIS-SUD 11



Centre d'Alembert

CENTRE D'ALEMBERT

Site Web : <http://centre-dalembert.u-psud.fr>

Colloque « Evaluations de la recherche »

Evaluer les recherches : acteurs, méthodes, fonctions et conséquences

13 et 14 Mai 2009

Mercredi 13 mai 2009, 10 heures

Sujet : *Histoire comparée de l'évaluation dans la gestion de la recherche publique et privée, fondamentale ou finalisée. Objets d'évaluation (individu, laboratoire, organisme) .Modalités et usages*

Evaluation scientifique et évaluation managériale des chercheurs et des institutions de recherche.

Bertrand BELLON.

Nous vivons une avalanche de réformes de notre système de recherche et d'enseignement supérieur, réclamée depuis trente ans par notre communauté, mais dont l'application se fait dans des conditions difficiles et discutables à de nombreux points de vue.

Nous nous limiterons ici à réfléchir sur le concept d'évaluation et son application aux activités d'enseignement et de recherche, ainsi qu'aux management des institutions des institutions qui en sont chargées.

Nous procéderons en quatre étapes :

- Quelques réflexions sur la nature de l'évaluation
- Les objets de l'évaluation,
- les évaluateurs,
- Et quelques modalités de mise en application

1- Evaluer :

En dépit des évolutions rapides observées depuis cinq ans, évaluer est une préoccupation ancienne dans le monde de la recherche.

Dans les faits, l'évaluation a été mise en oeuvre à des rythmes divers selon les établissements et selon les domaines. **Si dès 1945 le CNRS se dotait d'un comité national de la recherche scientifique** dont la mission était d'évaluer l'activité des laboratoires et des chercheurs, mais ce n'est que depuis quelques années que l'évaluation des chercheurs titulaires est devenue la règle dans l'ensemble des EPST, **L'INRETS a ainsi attendu 1999** pour faire fonctionner une commission d'évaluation de ses chercheurs. Le principe de l'évaluation systématique et de la liaison entre résultats-et-moyens est directement formatée par les concepts de la LOLF.

Mise en application ou non, l'évaluation de la recherche possède une forte légitimité :

- D'abord, *parce que celui qui fait de la recherche n'est pas celui qui la paie, et encore moins celui qui en bénéficiera.*
- Ensuite, *parce que la recherche répond à des critères particuliers du temps et de l'espace : je fais de la recherche aujourd'hui grâce aux découvertes de tous ceux qui sont passés avant moi, hier et avant-hier ; et pour des résultats qui bénéficieront à mes enfants et à mes petits enfants.*
- Et également une dimension particulière de l'espace. *Les frontières de la recherche ne sont pas nationales ; Elles sont cognitives. Les échanges se jouent des frontières. Mais il ne sont pas partagés par tout le monde pour autant. Le résultat de la recherche, la connaissance est à la fois un bien public et un bien qui n'est pas accessible à tous. C'est aussi un bien qui peut être emprunté, volé, ou dépassé ; et ceci, indépendamment de la volonté du chercheur.*

Pour la période actuelle, ces préoccupations sont renforcées par plusieurs facteurs :

- *La mondialisation des connaissances est devenue le socle des changements économiques et sociaux. La recherche et l'enseignement supérieurs sont mis aux avant-postes. Il en résulte une multiplication des partenaires et un besoin de comparer les résultats et les pratiques (benchmarking), même lorsqu'on se limite aux caricatures des classements)*
- *Les problèmes propres à la raréfaction des finances publiques renforce la rationalisation de la gestion des budgets et favorise naturellement les décisions ayant un effet à court terme au dépens des projets lointains naturellement incertains.*
- *Le raccourcissement généralisé de la conception du temps, qui est à la fois une caractéristique de toutes les époques modernes, renforcée par la crise d'adaptation de nos systèmes nationaux à la mondialisation.*

Dans ce contexte, la recherche apparaît, par nature ambitieuse et incertaine. Elle mobilise une grande masse de moyens humains, de moyens financiers, de comportements individuels et collectifs et d'opinions. Se pose le problème de l'articulation entre les exigences de court terme de plus en plus fortes et la structure de long terme du monde.

L'évaluation participe de la nature stratégique de la recherche. Elle est devenue une obligation, une exigence. Dans ce cadre, c'est une contrainte ; mais c'est également une opportunité et une occasion de multiples effets pervers.

L'évaluation est une exigence politique du 21^{ème} siècle.

Les besoins économiques sociaux de toutes les sociétés (d'abondance comme les nôtres ou de pénuries, comme des deux tiers de la planète) restent immenses. Dans les domaines de l'énergie, des matières premières, de la biodiversité, de la santé, des communications ... après tant d'année de course folle à la croissance, il est devenu indispensable de se préoccuper des objectifs durables, et non seulement des exigences de court terme. Il n'est plus possible de considérer le développement comme une accumulation de biens et services, indépendamment de la société que l'on veut laisser à nos enfants et à nos petits enfants.

L'évaluation est une exigence économique : Celui qui effectue la recherche n'est pas celui qui la paie, ni celui qui en bénéficiera.

Simultanément, l'objet de la recherche, la science, qui est traditionnellement pluriel et changeant, est devenu concurrentiel et ouvert au niveau mondial. La concurrence scientifique en est transformée

C'est une activité aux effets externes majeurs. Plus prosaïquement, l'activité de recherche et son financement étant, majoritairement de source publique, le chercheur est nécessairement responsable vis-à-vis de celui qui paie les impôts ; c'est-à-dire du corps social dans son ensemble et des pouvoirs politiques qui le représentent. (réponse cette exigence conduit à porter l'évaluation au niveau correspondant : régional, national, européen et mondial)

La recherche fait apparaître deux problèmes économiques : celui des relations entre coûts/bénéfices et celui du risque.

Plus le monde est ouvert, plus la concurrence, porte non seulement sur les coûts de productions (prix de la main d'œuvre, coûts des matières premières et des équipements), mais sur les facteurs « hors-prix » ; les facteurs qui incluent l'amélioration des biens et services existants ; l'innovation ; c'est-à-dire la recherche dans tous les domaines et ses applications.

Depuis 25 ans maintenant, il faut non seulement produire moins cher, mais apporter davantage pour le même prix. Les mondes de l'électronique, de la communication et de la santé sont d'excellentes illustrations de cette règle mondiale.

Mais tout ceci implique de larges détours de production mobilisant la recherche la plus fondamentale. Pour proposer de nouveaux services avant les concurrents, il faut préparer le terrain longtemps à l'avance ; prendre le risque de la recherche. C'est une des contradictions majeures avec les sanctions à court terme que procure le marché mondialisé.

Mais cette exigence économique reste très complexe. Elle ne se limite pas à la stratégie, et à la mise en œuvre de la recherche ; mais elle pose la question de l'appropriation et du contrôle des résultats de la recherche ; ou plutôt de son exploitation. Quand les systèmes nationaux dominaient il y avait une relative adéquation entre les états ou les sociétés nationales qui finançaient, qui décidaient, qui faisaient et qui bénéficiaient de la recherche. Les bénéficiaires externes restaient marginaux. Aujourd'hui ce n'est plus le cas ; c'est plutôt l'inverse qui se produit.

C'est une exigence éthique :

Or il n'existe pas de relation directe entre la recherche et ses résultats applicables.

Tout chercheur sait que le résultat de son travail ne sera pas, la plupart du temps, bon ou mauvais en lui-même. Il a pour mission de faire reculer le niveau de l'ignorance. C'est ici une course sans limite. Mais c'est l'usage des résultats de la recherche qui lui donnera toute sa valeur ; ou au moins toute sa justification. **L'atome** apporte des progrès fantastiques en matière de santé et d'énergie en même temps que dans des moyens de destruction massive et des risques beaucoup plus disséminés. La **génétique** pose des problèmes sur lesquels nous avons chacun, notre *idée personnelle*, mais qui reste largement ouvert et incertain au niveau social. (Réponse : Cette exigence est une raison majeure dans la nécessité d'évaluer les pratiques autant que les résultats)

En oeuvrant pour l'avancée de la connaissance (l'amélioration de la situation matérielle des hommes, valorisation ou formation), la recherche court toujours le risque de jouer les apprentis sorciers ; et la mutation de nos sociétés en vastes espaces médiatiques ne facilite pas les choses. Vouloir que la recherche serve à quelque chose introduit des responsabilités nouvelles.

En conclusion, l'évaluation est une nécessité : dans un monde ouvert et concurrentiel, la course aux financements et la course aux applications de la recherche, renforce la course à la connaissance et à la reconnaissance, qui inclut naturellement celle à l'appropriation, légitime ou non.

Dans cet espace incertain tous les acteurs de la Triple Hélice (de Etzkowitz & Leydesdorff, 2000), ont un besoin absolu de savoir où ils en sont : l'étudiant ou le chercheur doit pouvoir choisir dans quelle université aller ; l'industriel quel centre de recherche mobiliser, le politique quelle institution financer ; le citoyen ce qu'il advient des choix et des financements publics auquel il contribue ; le journaliste, ne pas s'appuyer uniquement sur le dernier événement susceptible de faire un scoop.

Si nous ne proposons pas, nous-mêmes, des démarches ouvertes et adaptées d'évaluation de notre activité, elles seront faites de l'extérieur, forcément de manière plus rapide, plus simpliste, et naturellement plus inacceptable.

Mais, dans l'environnement international actuel du financement de la recherche sur la base des résultats obtenus, l'évaluation prend une signification restreinte mais très puissante qu'il faut préciser et sur lequel il faut agir.

C'est aussi un risque (l'ombre de la caverne de Platon) : rien n'indique que le chercheur trouvera. Ni que celui qui trouve en sera récompensé, ou même reconnu. Ni que celui qui ne trouve pas, verra son travail utilisé par d'autres pour faire avancer le patrimoine cognitif collectif. Le chercheur restera le plus souvent inconnu ; même s'il est abondamment pillé. L'évaluation est ce processus qui créera une image spécifique de l'activité de recherche. Cette image a naturellement plus de force persuasive que la réalité elle-même.

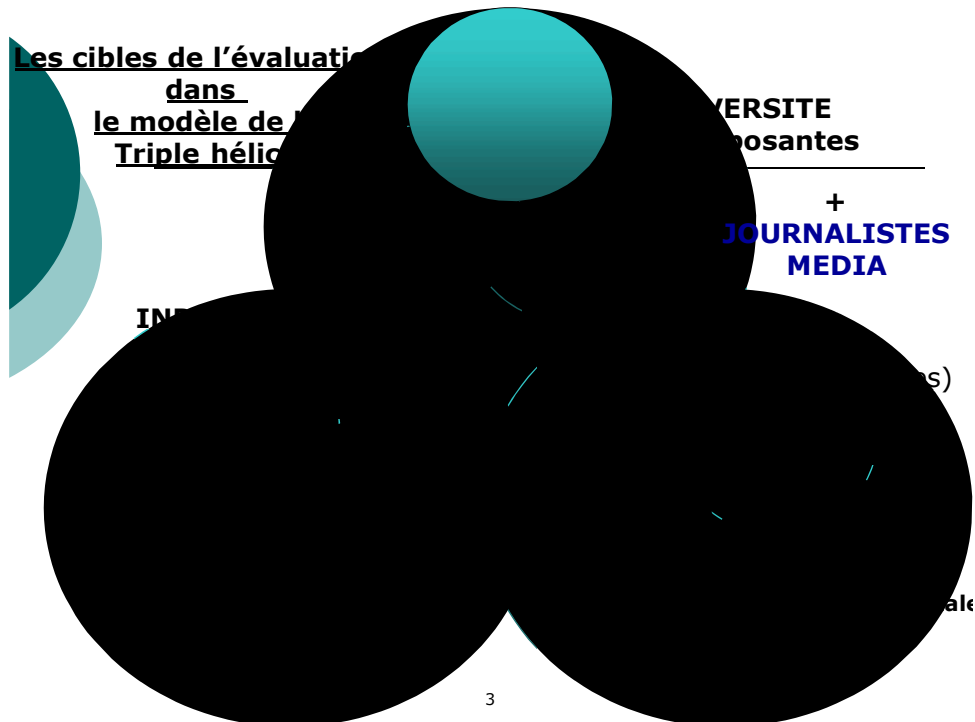
Aucune évaluation sera jamais objective, ni complète. Même conduite de manière systématique, l'évaluation est une discipline qui fait partie *des sciences sociales* et *non des sciences exactes* : elle est caractérisée par le fait que le degré de complexité qu'elle est susceptible de développer sera toujours largement en deçà au degré de complexité de l'objet à évaluer.

L'image, évidemment simplificatrice, sera non seulement reprise par les instances compétentes dont dépendent les centres de recherche ou les chercheurs, mais par le grand public, via le presse et les médias et se répandra, comme une trainée de poudre. Indépendamment de ses auteurs. Elle envahira l'opinion publique et tous les usagers individuels (nécessairement que partiellement informés) de l'appareil de recherche et d'enseignement supérieur (*l'important n'est pas ce qui est, mais ce que les gens croient* Taleyran).

C'est un objet de manipulation : On le voit tous les jours : En s'appuyant sur de grands principes comme le principe de précaution, ou en cherchant le scoop médiatique à tout prix, les informations chiffrées et ordonnées seront toujours l'objet de manipulations : il suffit de prendre la part des informations qui conforte la thèse défendue : ainsi, la grippe AH1N1 qui sévit actuellement dans le monde n'a fait que 54 morts à ce jour : C'est potentiellement le début d'une catastrophe, mais c'est aussi – à l'échelle d'une population de plus de 6 milliards d'habitants, une grippe banale, voire insignifiante. La réaction journalistique ont été la caricature de ce qui se passe habituellement. Un mouvement cumulatif nous a d'abord annoncé la pire catastrophe mondiale ; puis, les démentis se sont succédés et on a pratiquement oublié l'évènement. La réalité se trouve entre ces deux extrêmes : la situation pourrait être grave et il faut prendre des précautions. Mais il n'existe aucune crise avérée.

L'évaluation des phénomènes devient l'objet naturel de manipulations. L'évaluation va avoir une dimension comparative, même si elle n'a pas été conçue dans ce but. Il faut donc porter une attention particulière à :

- La nature de l'évaluation (des moyens mobilisés, du processus, des résultats...)
- A sa cible (les évalués, les financeurs, les usagers et bénéficiaires...)
- Le moment et la manière de dévoiler l'information
- Les conflits d'intérêts et l'identification des évaluateurs
- L'irrésistible tendance aux simplifications de l'utilisation des évaluations (rôle clé des médias ; Classement de Shanghai).



Réponse à ces limites : 1/ considérer les évaluations comme un des éclairages permettant de montrer l'état des lieux et les évolutions ; mais pas le seul ; 2/ ne pas appliquer mécaniquement de lien entre les moyens accordés à la recherche et les évaluations ; 3/ assurer aux évaluations une réelle **transparence** et aux évaluateurs une réelle **autonomie** par rapport aux objets de l'évaluation et aux intérêts en présence ; 4/ rendre les évaluations contradictoires et les agences d'évaluation concurrentes (assurer une évaluation des évaluateurs).

C'est enfin une opportunité :

Enfin, pour chaque acteur de la recherche et de l'enseignement supérieur l'évaluation est une opportunité.

D'abord au niveau de son processus

Le processus choisi pour conduire une évaluation est toujours plus important que son résultat. Il demande, d'abord, de la part de chaque chercheur ou chaque équipe, une attitude auto-réflexive qui commence avec le cahier de laboratoire et qui a son point culminant au moment de la visite du laboratoire par l'institution d'évaluation. Ces guides instaurent dans les pratiques un curseur permanent et dynamisent les comportements individuels et collectifs. Elles mettent en évidence cent innovations possibles et restées cachées.

Ensuite au niveau de ses résultats :

Les résultats permettent de se situer dans le temps (comment les pratiques ont-elles évolué) et dans l'espace, par rapport aux autres. Ils clarifient comment évolue la mission que l'on s'est fixé sur le long terme.

L'évaluation est donc un instrument de gestion et pas seulement un outil de décision. C'est aussi un outil fragile, à manipuler avec précaution.

Ceci dans un contexte où la découverte scientifique est un phénomène chaotique. Laisser la place à l'incertitude, à l'accumulation, aux expérimentations, aux erreurs (avoir droit à l'échec et tirer parti des échecs)

La démarche est complexe. L'évaluation doit tenir compte que, comme le disait Hubert Curien, l'horizon de la recherche c'est l'horizon du forestier, mais, en même temps que les collectivités scientifiques sont des collectivités comme les autres.

2- Les objets de l'évaluation :

Nous n'aborderons pas ici l'évaluation des travaux scientifiques individuels, notamment les systèmes d'évaluation par les pairs dans les comités de lecture (*dont parlera Philippe LAREDO dans un instant*). Nous n'aborderons pas non plus les classements (Shanghai, Times, Grandes Ecoles) qui limitent les évaluations à un nombre restreint de critères et gagnent en objectivité ce qu'ils perdent en signification.



11/05/2009

2

L'objectif est de distinguer ici deux types d'évaluation : celle des résultats de la recherche et celle du management des institutions de recherche. Les deux démarches visent les mêmes objectifs mais n'ont pas les mêmes fonctions et n'utilisent pas les mêmes moyens.

Il y a naturellement des points communs.

Vient immédiatement à l'esprit l'examen des résultats de la recherche et des moyens mobilisés. Puis des rapports entre les premiers et les seconds, dont nous allons parler ici. La démarche input/output ; rapport entre résultats et moyens, constitue un indicateur très global de l'efficacité d'un système (d'une équipe ou d'un chercheur individuel). Mais elle élimine tous les indicateurs non quantifiables, quand les processus de recherche et de management de la recherche sont extrêmement complexes en non mécaniques.

Output : naturellement on se focalise d'abord sur les articles et sur leurs facteurs d'impact.

On suppose que les articles bénéficient d'un jugement par les pairs. L'infométrie joue un rôle important dans l'évaluation de la recherche. Elle représente un énorme travail qui s'appuie sur des bases mondiales (Thomson ISI), européenne (Eurostat) ou nationale (OST).

Mais, articles et citations ne représentent qu'une partie (minoritaire dans plusieurs disciplines) de la recherche. Il faut y ajouter les livres, les conférences, les cours, les brevets, les projets et les formes multiples de diffusion de la connaissance.

Ici également des indicateurs existent, mais ils ont une portée extrêmement différente selon les disciplines. Ils font aussi l'objet de controverses qui ne sont pas seulement franco-françaises : en particulier au niveau des SHS (1/3 des effectifs scientifiques).

Inputs : se comptabilisent plus facilement sous forme de budgets. On est conduit à additionner les salaires, des dépenses d'investissement en immobilier et en équipements, les crédits de fonctionnement.

Mais les moyens nécessaires à la recherche diffèrent considérablement d'une discipline à l'autre.

Outre les salaires des personnels, la physique et les sciences expérimentales demandent des investissements en équipements considérables, tandis que les mathématiques et les sciences sociales mobilisent beaucoup moins de moyens en volume, tout en ayant une nécessité absolue de billets d'avions ou de matériel informatique.

Risques :

La démarche Input-Output a l'avantage de mettre en regard des données simples. Elle nécessite un accord assez fin sur la nature des inputs et celle des outputs à prendre en compte. Elles ont une signification immédiate. Pourtant, il faut déjà prendre des précautions techniques permettant de refléter **la dimension temporelle de la recherche**, autant que **les différences entre disciplines**. Il se passe d'abord fréquemment 3 ans entre une recherche et sa publication dans une revue à référé. Mais l'échelle de temps des chercheurs est très différente d'une discipline à l'autre. Celle de l'historien est la génération (l'output est le livre, comme dans la majeure partie des sciences humaines ; il est fréquent qu'on ne puisse soutenir sa thèse qu'après dix ans de travail ; souvent davantage) ; l'échelle de temps du physicien est plus courte (l'output est l'article) ; l'échelle de temps de l'informaticien ou du généticien est le mois ou l'année (l'output est la publication dans une conférence, le secret ou le brevet).

Enfin, le lien entre « performances » et « récompenses » est sujet à de nombreuses critiques. On soulignera par exemple, que les institutions avec de faibles moyens par tête ont souvent les plus faibles résultats par tête. Ici les « indicateurs de performance » sont ambigus : si vous ne mettez pas d'énergie dans le moteur, il n'est pas surprenant qu'il tourne moins vite. Et on sait que les raisonnements de cette nature ont des effets cumulatifs. Les meilleurs à un moment donné obtiennent la plus grosse part des moyens et deviennent encore meilleurs (naturellement ce raisonnement ne reflète qu'une part de la réalité).

L'attrait pour la simplicité des indicateurs présente un double risque : celui de passer son temps à évaluer et celui de tordre les comportements qui deviennent entièrement rivotés sur de tels indicateurs. Le mal est déjà fait : *on divise un article conséquent en cinq morceaux d'articles insignifiants ; on ne laisse plus de place à la recherche innovante qui risque fort de ne jamais passer dans des revues à référé ; on crée des effets cumulatifs au profit de quelques revues et de leurs équipes animatrices.*

Evaluation scientifique et évaluation gestionnaire.

La démarche input-output présente des intérêts et a des limites qui sont comparables pour les individus et pour les équipes et les organismes.

Il n'en est pas de même, lorsqu'on aborde le management interne des institutions. Les différences apparaissent au niveau des modes de gestion des moyens , et plus largement celui de la gouvernance (au sens général de direction, d'administration et de gestion).

Les indicateurs input-output font l'économie des modes de management de la recherche, ainsi que des finalités poursuivies.

Les modes opératoires commencent par les cahiers de laboratoire, les démarches qualité et les actions variées de normalisation des activités recherche, et de l'enseignement supérieur, actuellement en rapide développement. L'ensemble du système d'évaluation des pratiques est actuellement formaté par la LOLF. La LOLF répartit les crédits de chaque programme en actions ; elle introduit la stratégie de la performance ; elle conduit à la gestion des crédits au premier euro, donc introduit une comptabilité en coûts complets (comportant amortissements, provisions frais de gestion et dépenses fiscales) ; le tout en liant (ou du moins cherchant à le faire) l'octroi des crédits aux résultats des exercices précédents.

A ce titre, la gouvernance a une acception plus large que ce qu'il recouvre souvent. Gouverner, c'est en particulier développer :

- La stratégie des institutions,
- Les relations internationales
- Animer les processus de recherche, autant que d'innovation
- La gestion quotidienne des hommes et des moyens qui crée cette alchimie relationnelle qui conditionne la capacité créative individuelle et collective

L'évaluation des institutions comprend, au premier chef, celle de leur gestion. Ce n'est pas l'institution qui fait la recherche, ce sont les chercheurs et leurs équipes. L'évaluation gestionnaire se concentre sur le mode d'utilisation des moyens et leur capacité à susciter, faciliter, encourager la production d'enseignement et de recherche, à développer l'attractivité, à susciter le développement de capacités nouvelles, les capacités organisationnelles, la valorisation et les interactions (coopération effective nationale et internationale, disciplinaire et inter-disciplinaire).

Pour conclure, on peut avancer 4 remarques :

- l'évaluation de la recherche ne se limite pas à l'évaluation des frontières de la connaissance (la science qui se fait, la science qui se transmet, la science qui se fera demain)
- les médias de la connaissance sont variés et leur rôle évolue rapidement : articles et livres ; publications électroniques, brevets et droits d'auteurs ; conférences et débats ; cours et enseignement à distance...
- l'évaluation scientifique (produire mieux) est très différente de l'évaluation gestionnaire (gérer mieux)
- Enfin, les agents à évaluer ne se limitent pas aux chercheurs et aux organismes de recherche ; ils incluent, naturellement et fortement les politiques publiques correspondantes et les organismes publics chargés de les concevoir et les promouvoir. Elle concerne tout autant les évaluateurs.

3- Les évaluateurs

L'évaluation (et encore davantage lorsqu'elle se fonde sur des critères quantitatifs) accorde un pouvoir exceptionnel à l'évaluateur. Ses résultats bénéficient d'une présomption d'objectivité pour les acteurs concernés ; surtout, ils seront toujours interprétés et utilisés dans un sens univoque par les acteurs extérieurs.

Il est donc nécessaire de bien identifier qui sont les évaluateurs, quelle est leur légitimité ; leur mode de fonctionnement ; les évolutions de leur pratique ; et les processus de leur propre évaluation.

Ce ne sont pas les institutions d'évaluation qui manquent. Tous les pays ont mis en place des processus complexes en ce sens.

Dans les pays centralisés, les institutions se rencontrent le plus fréquemment aux échelons hiérarchiques supérieurs par rapport aux institutions qui font l'objet d'évaluation.

Dans les pays décentralisés, l'évaluation est plus fréquemment assurée par des instances indépendantes et davantage incluses dans des procédures contradictoires ou « contestables ».

Dans les deux cas, hormis le niveau des moyens attribués à de tels processus, les questions qui déterminent la compétence et l'efficacité des procédures sont celles **du degré d'indépendance et du niveau de transparence des activités**, c'est-à-dire du degré de contestabilité des institutions d'évaluation.

En France, l'essentiel des procédures d'évaluation répond à la structure centralisée du pays. Elles se situent, principalement, aux échelons administratifs supérieurs.

Les institutions les plus fréquentes se retrouvent dans les ministères. Le Ministère de l'Education Nationale a une **Direction de l'évaluation, de la prospective et de la performance (D.E.P.P.)** ; il entretient le corps des **inspecteurs généraux de l'éducation nationale (IGEN)** ; le **Comité national d'évaluation des établissements publics à caractère scientifique, culturel et professionnel (C.N.E.)** ; la **Direction de l'évaluation, de la prospective et de la performance (D.E.P.P.)**.

Plus généralement, les rapports de la **Cour des comptes** complètent l'évaluation de l'utilisation des fonds publics dans tous les domaines.

Tous **les grands corps de l'Etat** (assemblée nationale, Sénat, autres ministères) ont, également, leurs propres instances d'évaluation.

Par ailleurs, **les laboratoires de recherche et des instituts, et, à fortiori les grandes écoles**, ont leurs propres **processus qualité** concernant organisations, politiques et résultats.

Quelques organismes d'évaluation plus indépendants existent comme la **Commission des titres de l'ingénieur**.

Les récentes réformes tordent cette répartition des tâches en introduisant une nouvelle **culture de la performance et du résultat**. Cette culture favorise l'association entre évaluation et allocation des moyens. Elle a également pour conséquences de rapprocher évaluation scientifique et évaluation managériale. Ceci constitue le cadre de la dernière née des instances publiques et nationales, **l'AERES**.

Les modes d'évaluation par étapes du type AERES :

- Auto-évaluation par l'organisme concerné, pouvant durer trois à quatre mois
- Comité d'experts
- Visite sur place,
- Rapport de synthèse validé et transmis à l'organisme et à ses autorités de tutelle, avec réactions.
- Le tout rendu public.

Quelques idées issues des comparaisons internationales :

En Allemagne, les organismes publics et para-publics d'évaluation sont doublés de commissions indépendantes comme le souligne le récent (second) Rapport d'évaluation sur la recherche,

l'innovation et la puissance technologique de l'Allemagne (mars 2009). D'autres organismes comme le Centrum für Hochschulentwicklung / CHE existent et proposent notamment un « classement non classé » des universités de langue allemande sur des bases originales. En Allemagne, les agences d'accréditation sont des sociétés privées et opèrent au niveau des länder. La différence qui marquait le paysage universitaire allemand entre des universités de technologie (*Fachhochschulen*) tournées vers la pratique, d'un côté, et des universités tournées vers la science, de l'autre, est en cours de disparition.

Les États-Unis, représentent un système très décentralisé, couplé avec de forts moyens mobilisés à chacun des niveaux considérés.

Il n'y existe pas de coordination entre les financements de la recherche au niveau fédéral et au niveau des Etats. Les actions de recherche et d'enseignement supérieur sont à la fois largement soutenues par des fonds publics (niveau comparable à celui de la France) et par des fonds privés (niveau nettement supérieur). Les organismes d'évaluation privés sont plus importants que ceux du public et la concurrence domine. Les usagers de l'université et de la recherche sont des « clients » vers lesquels sont dirigés les différents niveaux d'évaluation. Les usagers de la science : étudiants, chercheurs, entreprises (ceux qui votent avec leurs pieds ou leurs carnets de chèques) sont les premiers évaluateurs ... ou les cibles des instances d'évaluation.

Dans ce contexte, l'analyse critique de tous les aspects de l'activité est bien davantage permanente. A preuve le fait que les autorités américaines (à l'intention de l'OCDE), font état d'inquiétudes que suscitent parfois les excès résultant d'une confiance exclusive dans les seules lois du marché ou une exacerbation de la compétition entre les institutions de recherche. Trop de concurrence marchande conduit dans plusieurs cas, à négliger la recherche fondamentale et à conforter un niveau de conformisme.

Le National Science Board (NSB) a également déploré le fait que la détermination des priorités budgétaires de la politique fédérale de recherche se repose davantage sur les résumés des évaluations que sur des analyses scientifiques suffisamment solides.

Comme le souligne une note récente de la Mission pour la science et la technologie de l'Ambassade de France à Washington, les États-Unis s'interrogent actuellement, aussi, sur l'efficacité de leur système de recherche et d'innovation. L'accent est mis sur une simplification des règles d'octroi des aides à la recherche et l'idée d'une contractualisation uniforme des relations entre les universités et les agences. Le *Program Assessment Rating Tool* (PART) a créé, en 2002, un cadre qui permet d'ajuster les moyens des agences en fonction de l'évaluation des résultats des projets.

Les Etats-Unis prennent également conscience :

- de leur dépendance à l'égard de la matière grise étrangère (restrictions dans la délivrance de visas d'entrée depuis le 11 septembre) quand 58 % des post doctorants employés aux États-Unis sont des résidents temporaires ;
- du risque de remise en cause de son leadership scientifique et technologique par la concurrence des pays émergents, surtout asiatiques.

Au Québec, Les systèmes d'évaluation sont nombreux et très ouverts. Ils ont des caractères mixtes à travers des systèmes fédéraux-provinciaux d'une part et des institutions indépendantes et concurrentielles. Les procédures d'évaluations sont nombreuses et mixtes. La conférence des présidents d'université (CREPUQ) s'est dotée de son propre organisme d'évaluation indépendant.

Le Royaume Uni est un système centralisé du point de vue de l'évaluation. Il a été un des premiers pays à adopter une politique de financement fondé sur les mérites. Dès 1986, le Higher education funding council of England a établi une évaluation globale qui a largement privilégié les

meilleurs résultats. Il a dû arrêter ses travaux quelques années après, notamment pour cause de manque de financement au profit des « institutions d'excellence » qu'il avait mis en exergue. Actuellement, les principaux sujets de débats concernent les inconvénients de la croissance excessive de la part du financement de la recherche sur projet.

Cette pratique entraîne :

- un manque de considération pour les sujets émergents ou les travaux fondamentaux de long terme ;
- un sous-investissement dans les infrastructures (un fond spécifique vient d'être créé pour remédier aux insuffisances constatées de ce point de vue).

Enfin, la coordination et la coopération entre les différentes structures de gouvernance et d'exécution de la recherche ont toujours fait l'objet d'une particulière attention des autorités britanniques et de nombreuses réformes ont été menées sur ce plan.

Par ailleurs, il existe de nombreuses organisations d'évaluations à caractère international, qui permettent de **comparer la performance des systèmes universitaires et de recherche** à travers le monde. On mentionnera entre autres :

- L'association internationale pour l'évaluation de l'efficacité dans le domaine scolaire (I.E.A.)
- L'organisation de coopération et de développement économiques (O.C.D.E.)
- L'Unesco ;
- Eurydice, le réseau d'information sur l'éducation en Europe.

L'obligation d'évaluer les évaluateurs :

Chaque instance d'évaluation a une tendance naturelle, et légitime, à justifier son travail (dans l'absolu et par rapport aux autres) et à étendre son champ de compétences (AERES par rapport aux titres des écoles d'ingénieurs). En même temps, chaque partenaire évalué a tendance à légitimer le niveau d'évaluation avec lequel elle a le plus de connivence.

Apparaît donc la question de la qualité, de la régularité et de la portée des évaluations externes dans le temps.

Evaluer les évaluateurs reste encore une question ouverte, quoi que chaque instance prétende naturellement se soumettre à un tel régulateur.

4- Modalités : Evaluer comment :

(Philippe Lazard parlera, dans un instant, de l'évaluation en tant que processus)

De nombreuses organisations sont maintenant spécialisées dans l'évaluation. Les techniques sont très diverses et nous ne ferons pas le tour de la question ici. Par contre deux remarques préliminaires méritent d'être faites : considérer l'évaluation comme un processus et non comme un résultat ; et, d'autre part, distinguer l'évaluation des différentes formes de classements.

L'évaluation est d'abord un processus, pas un résultat

L'évaluation cherche, d'abord, à s'appuyer sur des critères quantitatifs, ayant une certaine objectivité. Mais elle part d'une « auto-évaluation », et repose sur un processus d'interactions entre l'évaluateur et l'évalué. Un élément majeur est que (comme le fait l'AERES ou le CNRS) les procédures d'évaluation conduisent à faire se rencontrer les différents acteurs, à établir un dialogue, à comparer avec les situations analogues, à visiter les établissements et les équipes. L'évaluation est associée à l'idée qu'il existe, dans n'importe quelle organisation 20 à 30% de gisements d'efficacité inemployés. Ainsi, l'évaluation peut être envisagée comme une démarche globale destinée à identifier ce qui est et à l'améliorer, quelque soit le niveau atteint.

A côté des innovations incrémentales, les évaluations sont un outil de la mise en oeuvre des objectifs de la planification stratégique. Certains EPST (l'INRA notamment) ont d'ailleurs créé une structure rattachée à leur direction générale et chargée de suivre l'ensemble des évaluations. En ce qui concerne les personnels, l'évaluation a d'abord été conçue à l'intention des chercheurs. Bien que dans quelques établissements elle ait également été appliquée à certaines catégories d'ingénieurs, les personnels ITA étaient, en règle générale, soumis aux procédures d'appréciation prévues par le statut général des fonctionnaires. La réforme des statuts spécifiques des personnels des EPST intervenue en février 2002 a modifié cette situation en prévoyant une généralisation de l'évaluation à l'ensemble des personnels.

Evaluations et classements :

J'ai déjà montré à plusieurs occasions que, dans le contexte actuel, la seule réponse au classement de Shanghai n'était pas de le refuser mais d'en faire d'autres, sur d'autres bases, mais avec la même ambition planétaire.

Il faut dire ici, qu'en dépit des critiques faites à ce classement, qui a quelques fondements inattaquables, la culture de l'évaluation en France est profondément marquée par celle des classements (de l'école primaire à l'ENA ou Polytechnique ; de la sortie de l'école à la fiche nécrologique) comparée à celle des pays anglo-saxons (les dernières réalisations l'emportent sur les pedigrees).

Aux US, la réputation est un élément central de l'évaluation, parce que le système est tiré par le client (l'étudiant qui s'inscrit à l'université, le chercheur qui candidate, l'entreprise qui finance la recherche ou la formation).

En France, l'évaluation a une dimension beaucoup plus institutionnelle ; elle cherche à identifier dans quelle mesure la production scientifique se rapproche de standards validés au niveau national.

Ainsi, contrairement aux apparences et au discours égalitaire national, le système français est très familier avec les notes et les classements.

Au lieu de démarches qualitatives et progressives, le système français ordonne, catégorise et classe. Cette démarche a longtemps favorisé un grand écart entre des ordres de classement et l'absence totale d'évaluation. On aime à mettre des notes ; à indiquer des ordres ; voir, à lister des normes. Les notes à l'école primaire, secondaire, comme aux concours d'entrée et de sortie. Les pratiques de sélection des grandes écoles, comme les niveaux de spécialisation disciplinaire (difficultés de l'Université de Technologie de Compiègne lorsque Denielou a prétendu produire des « ingénieurs philosophes ») sont particulièrement inscrits dans nos traditions.

Les classements, qui prennent souvent en compte des indicateurs complexes, ne permettront jamais de répondre aux besoins de surveillance (monitoring), de contrôle, de fiabilité des institutions ; Ils ne permettent pas une meilleure gouvernance et ne sont pas en mesure de donner des indications sur les capacités propres de nos institutions ; notamment celles

- **D'autonomie et de capacités stratégiques**
- **D'attractivité**
- **De capacités de spécialisation et de différenciation**
- **D'implantation dans l'environnement local**

Il est clair qu'évaluer n'est pas classer. Les classements supposent une identification claire d'acteurs autonomes et une capacité à les positionner les uns par rapport aux autres, dans un ordre de mérite. Une raison à cela : la recherche (Calon-Latour) est un processus complexe où tout se tient : sauf de très rares exceptions fortement médiatisées, le génie ne peut s'épanouir que s'il se situe dans un milieu.

Mais l'évaluation sera toujours partielle et insuffisante : la seule solution viable est ici, comme pour les classements, d'en envisager plusieurs et de les considérer dans leur contexte à la fois relatif et interactif.

Réf.

- Evaluation : http://www.unicaen.fr/services/puc/article.php?id_article=732
- <http://indicasciences.veille.inist.fr/spip.php?article46>
- 2004 : Indicateurs utiles au management de la recherche issus d'un corpus multi-sources
Par MULTON (Jean-Louis) ; BRANCA-LACOMBE (Geneviève) ; DOUSSET (Bernard) ; DAGUILLANES
- Rapport « priorité à la recherche » 1999, Pierre Cohen – Jean-Yves Le Deaut
Documentation Française
- Pacte pour la recherche de 2006
- Loi sur les libertés et responsabilités des universités du 10 août 2007