

Pour une nouvelle démarche d'évaluation des laboratoires de recherche, des enseignants-chercheurs et des chercheurs en économie

Document collectif

Coordination : Gabriel Colletis et Thomas Lamarche¹

La Commission évaluation de l'AFEP a regroupé une trentaine d'enseignants-chercheurs et de chercheurs qui se sont organisés en trois sous-groupes : l'un consacré à la question de l'évaluation des laboratoires de recherche, l'autre à celle de l'évaluation des enseignants-chercheurs, le troisième à la question des publications.

Ces trois sous-groupes ont travaillé en parallèle et de façon concertée pour produire le présent document. La méthode de travail a donc été itérative et a favorisé le débat.

Au final, les contributeurs ont convergé sur l'essentiel et ont maintenu certaines différences d'appréciation sur tel ou tel aspect de l'analyse et des préconisations (en particulier sur l'AERES).

Le présent document se compose de cinq parties :

- **La première partie (I)** pose la question générale de l'évaluation dans le contexte actuel en examinant les différents aspects des pratiques actuelles, et que l'AFEP entend contester. Au final, apparaît clairement la nécessité d'une clarification de l'intention de l'évaluation. À l'opposé de l'actuelle évaluation sanction, que l'on peut suspecter d'introduire de sérieux biais, l'AFEP propose une évaluation professionnelle visant à promouvoir la qualité et le pluralisme des approches, des méthodes et des objets de recherche.
- **La deuxième partie (II)** analyse les conditions d'évaluation des laboratoires de recherche. La question fondamentale est celle de leur cohérence avec les missions que la société assigne à ces laboratoires, missions parmi lesquelles l'élaboration de questions de recherche ayant une finalité d'avancée des connaissances et/ou de production de savoirs utiles est essentielle. Le corollaire de cette question est la façon dont on se représente ce qu'est un laboratoire.

Dominant actuellement la représentation d'un laboratoire comme somme de chercheurs produisant et des modalités d'évaluation reflétant cette conception. L'AFEP entend rappeler que les missions fondamentales des laboratoires ne peuvent être réalisées qu'à la condition de considérer qu'un laboratoire est un collectif de chercheurs rassemblés autour d'un projet scientifique porté par des valeurs partagées.

¹ La liste des contributeurs des parties II à IV est indiquée dans les parties correspondantes. L'ensemble des contributeurs est donné page 6.

Les modalités d'évaluation des laboratoires que l'AFEP entend promouvoir sont en cohérence avec la dimension collective, le rôle social et la qualité scientifique attestée par un projet des laboratoires concernés.

- **La troisième partie du présent document (III)** concerne l'évaluation des enseignants-chercheurs. Dans l'actuelle pratique d'évaluation, les enseignants-chercheurs (EC) sont pour l'essentiel évalués en fonction de leur activité de publication. Si ce n'est pas le rang des publications mais leur nombre qui détermine la qualité de chercheur "publiant" ou "produisant"² (2 pour un EC, 4 pour un chercheur), l'appréciation de la "qualité" de l'EC dépend en revanche du rang des publications issu de la liste de la section 37 du Cnrs. Les conséquences néfastes de cette façon d'évaluer l'activité des enseignants-chercheurs sont nombreuses et considérables, notamment pour ce qui est de l'innovation scientifique. L'AFEP propose de reconsidérer profondément le mode d'évaluation des enseignants-chercheurs en reconnaissant la diversité de leurs missions et celle de leurs pratiques. Plutôt que de les considérer comme des individus isolés et calculateurs, intéressés de façon principale par l'optimisation de leur carrière et leur accès aux primes, l'AFEP entend promouvoir la liberté du chercheur et son autonomie dans le cadre de collectifs orientés par le souci d'une progression de chacun et de tous, au service de la société.
- **La quatrième partie (IV)** rassemble les préconisations de l'AFEP dans la perspective d'une refonte très profonde de la démarche d'évaluation. Cette partie propose tout d'abord des préconisations à caractère général puis des préconisations aux deux niveaux appréhendés dans ce document : celui des laboratoires, celui des enseignants-chercheurs. L'ensemble de ces préconisations résulte des débats que la commission a eus pendant le temps de son existence, des analyses qui ont été produites, de l'examen aussi de travaux disponibles portés à la connaissance de la commission par ses membres. Un espace spécifique a été créé sur le site de l'AFEP afin de déposer certains de ces travaux ou documents.
- **La cinquième et dernière partie du présent document (V)** rassemble sous la forme d'annexes des extraits de documents qui ont été utilisés pour la production de ce travail.

² La non-distinction "publiant" ou "produisant" est édifiante. Un chercheur est désormais qualifié de "produisant" uniquement s'il publie un nombre donné d'articles référencés dans la liste AERES. Sa qualité de "produisant" sera ensuite qualitativement évaluée par recours à la liste hiérarchisée de la section 37 du CNRS. Tout autre activité de publication (les livres, en particulier) et, plus largement, son activité de recherche non exprimée par des articles (comme la participation à des opérations de recherche) sont, de facto, dévalorisées.

Sommaire

Liste des contributeurs 6

I. La question générale de l'évaluation : les pratiques actuelles et la nécessité de clarifier l'intention de l'évaluation 7

<i>I.1. Le contexte.....</i>	<i>7</i>
<i>I.2. Sur la question même de l'évaluation.....</i>	<i>8</i>
<i>I.3. La mesure</i>	<i>9</i>
<i>I.4. Conformisme et danger pour la démocratie</i>	<i>11</i>
<i>I.5. Une perte de pouvoir des instances de la profession.....</i>	<i>13</i>
<i>I.6. Le pouvoir et la norme</i>	<i>13</i>
<i>I.7. Le question-clé de la diversité.....</i>	<i>14</i>
<i>I.8. L'évaluation comme "système".....</i>	<i>17</i>
<i>I.9. Les effets du système d'évaluation sur chacune de ses composantes et leurs liens.....</i>	<i>18</i>
I.9.1 Les enseignants-chercheurs	18
I.9.2 Les laboratoires.....	19
I.9.3 Les liens entre l'évaluation des enseignants-chercheurs, celle des laboratoires et celle des Universités.....	19
<i>I.10. Au final, la nécessaire clarification de l'intention stratégique</i>	<i>20</i>

II. L'évaluation des laboratoires de recherche 22

<i>II.1. La réaffirmation du rôle central et de la diversité des missions des laboratoires.....</i>	<i>22</i>
<i>II.2. Clarifications quant à l'intention stratégique de l'évaluation des laboratoires.....</i>	<i>23</i>
<i>II.3. Les principaux principes de l'évaluation des laboratoires que l'AFEP défend</i>	<i>24</i>
<i>II.4. Les domaines et les méthodes de l'évaluation à finalité de classement des laboratoires (les pratiques actuelles)</i>	<i>26</i>
<i>II.5. Constats concernant l'AERES et position de l'AFEP.....</i>	<i>27</i>
Premier enseignement : une surpondération confirmée de l'activité de publication.....	27
Deuxième enseignement : une corrélation curieuse entre le taux de producteurs et la "qualité de la production scientifique"	27
Troisième enseignement : la variabilité de l'évaluation d'une vague à l'autre	28
Quatrième enseignement : des éléments de ratios défavorables aux laboratoires en développement.....	28
Cinquième enseignement : une procédure d'évaluation entraînant des coûts rédhibitoires pour les petits laboratoires.....	28
Sixième enseignement : des chercheurs au sein des laboratoires devant des dilemmes difficiles	29
Septième enseignement : une contestation à la fois large et profonde des évaluations des laboratoires par les laboratoires concernés	29
Enseignement hors analyse statistique : une évaluation à géométrie variable.....	29
<i>II.6. Préconisations générales</i>	<i>32</i>
<i>II.7. Préconisations concernant l'AERES dans l'éventualité de son maintien.....</i>	<i>33</i>

III. L'évaluation des EC	34
<i>III.1. Le primat de la recherche sur les autres missions</i>	<i>35</i>
III.1.1 Le rôle décisif de deux niveaux de listes qui trient et hiérarchisent	35
III.1.2 Resserrement autour d'une élite éloignée de la communauté.....	37
<i>III.2. La carrière : étranglement autour d'une seule mission, évaluée de façon uniforme....</i>	<i>38</i>
III.2.1 La thèse.....	38
III.2.2 Les qualifications et le recrutement de MCF.....	39
III.2.3 De MCF à Professeur.....	40
III.2.4 Recrutement des chargés de recherche et passage de chargé de recherche à directeur de recherche.....	41
III.2.5 Les promotions (Hors classe)	42
<i>III.3. Les propositions de l'AFEP</i>	<i>45</i>
IV. Synthèse et préconisations.....	47
<i>IV.1. Pour une évaluation professionnelle et non une évaluation sanction.....</i>	<i>47</i>
IV.1.1 S'assurer a minima du respect des règles et des procédures	47
IV.1.2 Accompagner et non vérifier la conformité à une norme	48
IV.1.3 Pour une évaluation fondée sur le contenu réel du travail d'enseignement et de recherche et non sur un régime d'évaluation quantitative.....	49
<i>IV.2. Préconisations concernant l'évaluation des laboratoires de recherche</i>	<i>50</i>
IV.2.1 La prise en compte la dimension longue du temps de la recherche.....	50
IV.2.2 L'évaluation par les pairs	51
IV.2.3 Le principe essentiel de lecture des travaux	51
IV.2.4 L'usage à modérer fortement des indicateurs bibliométriques.....	52
IV.2.5 Le soutien de thématiques de recherche innovantes.....	52
IV.2.6 La prise en compte des apports méthodologiques	52
IV.2.7 Un soutien privilégiant les laboratoires en difficulté ou en démarrage	52
<i>IV.3. Préconisations concernant l'évaluation des enseignants-chercheurs.....</i>	<i>53</i>
IV.3.1 L'évaluation aujourd'hui : l'évaluation sanction	54
IV.3.2 L'évaluation professionnelle	55
V. Annexes	58
<i>Annexe 1</i>	<i>59</i>
L'AERES et le mythe de la "masse critique"	59
<i>Annexe 2</i>	<i>61</i>
Interprétation des notations de l'échelle de notation multicritères	61
<i>Annexe 3</i>	<i>63</i>
L'évaluation contre la science ; productivité académique vs contribution scientifique : le cas de l'économie	63
<i>Annexe 4</i>	<i>72</i>
La loi relative aux libertés et responsabilités des Universités : essai de mise en perspective	72
<i>Annexe 5</i>	<i>73</i>
Du travail à l'action publique : quand les dispositifs d'évaluation prennent le pouvoir.....	73
<i>Annexe 6</i>	<i>75</i>
Réflexions croisées sur la mesure et l'incertitude	75
<i>Contribution de François Vatin : L'incertitude du monde social et la raison métrologique : une approche pragmatiste.....</i>	<i>75</i>

<i>Contribution d'Alain Caillé : Petit plaidoyer pour une incertitude modérée et une mesure mesurée</i>	76
<i>Contribution d'Olivier Favereau : L'incertitude sans mesure</i>	77
<i>Annexe 7</i>	79
Revue qui compte, revue qu'on compte : produire des classements en économie et gestion.	79
<i>Annexe 8</i>	82
External review report of the French Evaluation Agency for Research and Higher Education (AERES) with regard to the Standards and Guidelines for Quality Assurance in the European Higher Education Area, May 2010.....	82
<i>Annexe 9</i>	86
Le danger des mauvais usages des indicateurs	86
<i>Annexe 10</i>	87
Réflexion sur l'évaluation et les listes de revues. L'exemple de l'économie et de la gestion.....	87
<i>Annexe 11</i>	90
Une étude comparative internationale des bases de données des revues scientifiques en sciences humaines et sociales (SHS)	90
<i>Annexe 12</i>	91
Rapport de l'Académie des Sciences Du bon usage de la bibliométrie pour l'évaluation individuelle des chercheurs.....	91
<i>Annexe 13</i>	93
Quelques enseignements statistiques de l'évaluation des laboratoires par l'AERES	93
<i>Annexe 14</i>	99
Pour un débat sur les représentations en économie	99
<i>Annexe 15</i>	103
Une nouvelle catégorie de revues "multidisciplinaires ou d'autres disciplines" (MAD).....	103

Liste des contributeurs

Pierre Alary

Florence Audier

Jean-Pierre Chanteau

Hervé Charmettant

Gabriel Colletis

Antonella Corsani

Ai-Thu Dang

Florence Jany-Catrice

Thomas Lamarche

Catherine Laurent

Philippe Méhaut

Sandrine Michel

Martino Nieddu

Claire Pignol

Marianne Rubinstein

Christian du Tertre

Serge Walery.

Le texte produit a été amendé et adopté par le Conseil d'administration de l'AFEP.

I. La question générale de l'évaluation : les pratiques actuelles et la nécessité de clarifier l'intention de l'évaluation

I.1. Le contexte

En France, le service public d'enseignement supérieur et de recherche est l'objet de profondes transformations. Deux transformations principales méritent d'être rappelées. La première a trait au processus de massification de l'enseignement supérieur, la seconde concerne l'évolution profonde des conditions de production sociale des activités d'enseignement et de recherche. Cette seconde transformation interroge directement la nature de l'évaluation de ces activités.

Nous examinons succinctement ces deux transformations.

La massification

La France a franchi le seuil de son 1^{er} million d'étudiants à la fin des années 1970 et le seuil de son second au milieu des années 1990. Aujourd'hui, un peu plus de la moitié des étudiants inscrits dans le supérieur sont à l'Université. Cette massification de l'enseignement supérieur n'a pas été anticipée. Et ce au moins à deux niveaux. Tout d'abord au niveau des moyens, elle est intervenue à moyens constants ou légèrement croissants. L'Université et, par conséquent, les enseignants-chercheurs réalisent leurs missions dans un contexte dégradé. Ces difficultés trouvent en grande partie leur origine dans une course effrénée à la mise en concurrence des formations du supérieur dont l'Université pâtit. Ensuite au niveau des pratiques pédagogiques. Le modèle "cours en amphi – Travaux Dirigés" continue à prévaloir là où la massification suppose d'être attentif à la diversification des publics et de leurs besoins. Les étudiants qui parviennent à réussir dans cette organisation très sélective, souvent les mieux dotés socialement, intégreront les formations plus prometteuses des masters. D'autres les intégreront à partir de cursus protégés, réalisés hors de l'Université proprement dite.

La massification impensée de l'enseignement supérieur conduit à mettre le focus sur les promesses non tenues de l'accès au supérieur pour accéder à de meilleurs destins professionnels. Et simultanément, à faire des enseignants-chercheurs les boucs émissaires d'un échec jugé, à juste titre, socialement inacceptable. La mise en place des dispositifs d'évaluation des personnels vise ici à réallouer le temps de travail du plus grand nombre pour encadrer les étudiants, laissant à une petite minorité soigneusement sélectionnée le temps de la recherche. La volonté de moduler les services est avant tout un habillage de la pénurie. Mais, elle maintient en l'état le problème majeur de ne pas articuler l'enseignement à la recherche, vecteur de l'amélioration de l'efficacité du travail des enseignants-chercheurs mais aussi des formations supérieures utiles à une production sociale elle aussi en transformation, de plus en plus étroitement dépendante de compétences professionnelles issues de la formation par la recherche.

La production sociale

En dépit de ces difficultés, l'enseignement supérieur produit des diplômés qui répondent aux besoins de la production sociale. Son efficacité nécessite des savoirs plus complexes et des compétences plus adaptatives, que la formation par la recherche contribue à produire. Ce lien avec l'expérience professionnelle exerce, en retour, un effet de demande sociale sur l'évolution de l'enseignement supérieur.

La place de l'Université dans la société est en effet elle aussi en mouvement. Sa sanctuarisation appartient à une époque définitivement révolue. La formation et le contrôle des savoirs par une minorité ou bien par le plus grand nombre sont bien l'enjeu majeur de la production mais aussi plus largement du vivre ensemble dans une communauté humaine profondément transformée par le rôle crucial des savoirs.

Dans ce cadre, l'introduction de l'évaluation dans tous les compartiments des activités des institutions de recherche et de formation est le résultat de deux tensions contradictoires.

D'un côté, l'évaluation des formes que prend la coopération dans le travail, est une aspiration des personnels. Elle relève de la volonté de faire reconnaître un travail quotidien dont la charge ne cesse de s'alourdir, de se complexifier voire de se disperser, de carrières passionnantes mais plus difficiles et au sein desquelles les efforts réalisés par les individus se heurtent à des perspectives d'évolution de plus en plus confiscatoires. Au bord de la rupture dans l'exercice des métiers, l'évaluation est ici revendiquée pour donner à voir un travail réel mais invisible, les équilibres instables que crée le manque de moyens mais aussi pour permettre à chacun de se projeter dans un meilleur exercice de son métier au service de ses missions.

De l'autre côté, l'évaluation (des "personnes") est un outil de gestion visant à élever la productivité (apparente) des personnels dans un contexte de pénurie durable de moyens. Ainsi, l'évaluation se veut l'énoncé normatif d'une vérification et la légitimation d'une sanction. Chacun ici doit rendre des comptes permettant de trier entre ceux qui se conforment aux règles et les autres. La fabrique des règles privilégie une mission, la recherche, réduite de façon caricaturale à son aspect dénombrable. Décidées à peu, les règles expriment l'exercice d'un pouvoir d'auto-sélection et sont compatibles avec une certaine adaptabilité. Se développe ainsi une véritable ingénierie du soupçon prospérant dans un univers professionnel se nourrissant de la division des personnels.

Dans les deux cas, l'évaluation et la question des moyens font l'objet d'âpres discussions. À juste titre. De nouveaux équilibres sont sans cesse reconstruits. Il est vrai que le sujet est sensible au rapport de forces.

Cependant, faire émerger de nouveaux dispositifs d'évaluation, suppose d'élargir la perspective et donc de penser les transformations affectant l'enseignement supérieur et la recherche et leurs productions dans le contexte plus large de leur insertion (contradictoire) dans la société.

I.2. Sur la question même de l'évaluation

Évaluer semble être la réponse à des injonctions qui ne convergent pas spontanément : il s'agit tout à la fois d'assurer la *qualité* de la dépense publique à un moment où celle-ci est soumise à de très fortes tensions, de répondre à un besoin très fort de démocratie dans le fonctionnement et la vie-même des institutions. On s'aperçoit, en effet, que le débat sur les indicateurs (dont on pourrait penser en première analyse qu'il concerne spécifiquement le fonctionnement des institutions) finit fréquemment par interroger le système d'objectifs et, au-delà, les missions assignées aux institutions³.

³ M. Salles (Présentation du dossier "Représentations, modèles et normes pour l'entreprise", *Revue Droit et Société*, pp 21-26, n° 65 / 2007) souligne que les indicateurs sont des normes mais que celles-ci ne peuvent être

Évaluer est donc un exercice difficile mais nécessaire, réalisé souvent dans des conditions paradoxales. Le paradoxe se résume le plus souvent dans le repérage d'une tension : la légitimité persistante du principe d'évaluation et l'illégitimité récurrente des dispositifs d'évaluation (Annexe 5, M.-C. Bureau).

Dans un contexte où les classements de revues se multipliaient au profit de la mise en place de la bibliométrie, la section 37 du CNRS, composée d'économistes et de gestionnaires, s'est lancée dans la production de son propre classement. Poursuivant un travail non abouti durant la mandature 1997-2000, la section a formé en 2001, sous l'impulsion du président de la nouvelle mandature, une "commission revues" dont l'objectif était sans équivoque : produire un classement à usage interne, pour l'évaluation des candidats aux concours et des unités de recherche, prenant donc la forme d'un "outil d'aide à la décision plus que celle d'un outil absolu de classement qui donne des résultats aberrants et ôte tout besoin d'une instance d'évaluation et de jugement scientifique » (Annexe 7, Pontille et Torny, 2010).

Le classement de revues, par exemple, a été intégré par les grandes écoles de commerce en tant qu'outil d'évaluation de la recherche. Ce qui a(vait) été conçu comme une aide au jugement par les pairs est donc devenu un outil de calcul direct pour une institution, aussi concret que le taux d'encadrement qui rapporte le nombre d'étudiants aux enseignants permanents (Annexe 7, Pontille et Torny, 2010).

Les conditions matérielles de l'évaluation ont cependant un effet direct sur la nature de l'évaluation elle-même qu'il convient de ne pas négliger. "La montée en puissance du "régime d'évaluation quantitative "(REQ) est (...) le fruit de la réalité matérielle à laquelle se trouvent confrontés les évaluateurs. Les membres de ces institutions, contraints de rationaliser leur travail d'évaluation, sont naturellement conduits à recourir à des indicateurs quantitatifs" (Annexe 3, Wallery, 2011).

1.3. La mesure

Le débat sur la mesure, ses limites, voire son existence-même est essentiel.

Si le caractère réducteur de la mesure est souvent admis, certains estiment que la mesure est néanmoins incontournable dans la conduite des exercices d'évaluation.

"La mesure est réductrice. C'est là sa fonction même. On ne produit de l'information qu'en en détruisant. Il s'agit de réduire l'infini chatoisement du réel à un cadre métrologique qui en rend compte de façon simplifiée. Confrontée au réel, la mesure montrera elle-même ses limites, appellera à son renversement paradigmatique, quand d'autres mesures plus fines seront disponibles. Une mesure totalement arbitraire resterait préférable à l'absence totale de mesure" (Annexe 6 Vatin, in Vatin, Caillé et Favereau, 2010).

Admettre l'utilité de la mesure ne signifie pas cependant qu'une mesure unique, une "note" donc, soit une perspective envisageable.

"En tout état de cause, précisément parce que la valeur relative des divers types de savoirs est indéterminable, parce que leurs composantes sont nécessairement plurielles, l'idée même

comprises qu'en référence à des modèles eux-mêmes fondés sur des représentations. Le débat sur les missions possibles d'une institution se situe au niveau des représentations. Celui sur ses objectifs repose sur le niveau des modèles. Les objectifs à leur tour sont déclinables en termes de normes et s'expriment par des jeux d'indicateurs.

qu'il serait possible d'arriver, fût-ce approximativement, à une bonne mesure unique, se révèle totalement intenable" (Annexe 6 Caillé, in Vatin, Caillé et Favereau, 2010).

Mais il est possible d'aller plus loin dans la critique de la mesure et des indicateurs en mettant en cause ceux-ci sur une base plus radicale.

"La critique des indicateurs, aussi utile soit-elle, ne saurait se substituer à celle de l'évaluation par indicateurs (...). L'AERES et l'ANR, par leur composition et leurs critères d'évaluation, constituent, à ce jour, l'aboutissement de cette logique : une évaluation sans lecture ni appréciation des travaux" (Annexe 3, Wallery, 2011).

Si les critères d'évaluation de l'AERES (voir annexe 2) doivent retenir l'attention, le langage employé, de type managérial, également. On relève des termes comme : leadership, excellence, performance, prix, visibilité internationale, compétitivité, ambition...

Ce qui est excellent est "international", ce qui ne l'est pas ou l'est moins est "national" ou, pire, "local"...

Parler de *politique* suppose d'inscrire les choix collectifs dans un projet de société publiquement exprimé et justifié ; lui substituer celui de *management* contient implicitement l'idée qu'il n'y a pas de débat possible – ou nécessaire – sur les fins, mais seulement une expertise à développer sur les moyens (Annexe 4, Vinokur, 2008).

Du point de vue du manager de la recherche, de l'évaluateur ou parfois du chercheur, le recours systématique, voire exclusif, aux indicateurs quantitatifs et aux classements qui en découlent, présente un quadruple intérêt : simplifier et rationaliser le travail d'évaluation, circonscrire les débats et controverses, donner une apparence d'objectivité aux classements obtenus, et "cardinaliser" les objectifs de la compétition des chercheurs (Annexe 3, Wallery, 2011).

Le dispositif d'évaluation de la recherche en économie apparaît comme un "*dispositif global*", au sens de Michel Foucault, qui assure le pouvoir de la norme sur la population des individus, au moyen d'un jeu incessant de visibilité, diagnostic et classement.

"On entend par "dispositif" tout ce qui a d'une manière ou d'une autre la capacité de capturer, d'orienter, de déterminer, d'intercepter, de modeler, de contrôler et d'assurer les gestes, les conduites, les opinions et les discours des êtres vivants. (...) Le dispositif est donc, avant tout, une machine qui produit des subjectivations et c'est par quoi il est aussi une machine de gouvernement". Là réside le principal pouvoir de la batterie d'indicateurs, d'outils et de protocoles qui constituent la machinerie de l'évaluation : elle transforme nos manières de sentir, de raisonner et de réagir même lorsque les dispositifs eux-mêmes sont peu opérants et peu légitimes" (Annexe 5, M.-C. Bureau).

La place tenue par les indicateurs aujourd'hui ne peut qu'interroger. Tout se passe comme si ceux-ci étaient devenus la fin, l'objectif ultime alors qu'ils ne devraient être que des moyens, des objectifs intermédiaires au service des missions assignées au système de recherche : faire progresser les connaissances, produire des savoirs utiles. Il n'est donc pas exagéré de parler de "détournement" : le moyen (l'indicateur) se substituant aux finalités (les progrès de la recherche, notamment par la prise en compte des attentes et des besoins de la société dans la formulation même des questions de recherche ainsi que par la diffusion des résultats).

Dans ce dispositif, la *bibliométrie* joue une place essentielle. La critique qui semble la plus complète sur cette question semble venir d'Yves Gingas, statisticien québécois, qui a écrit un article absolument définitif intitulé "Du mauvais usage de faux indicateurs", paru en 2008

dans *La revue d'histoire moderne et contemporaine*, et repris dans la Vie de la Recherche Scientifique de septembre 2008.

Il montre, entre autres, le détournement complet de la bibliométrie, initialement conçue par une petite communauté de bibliothécaires, sociologues, statisticiens qui étudiaient les transformations de la recherche à travers les propriétés des publications et les références qu'elles contiennent. Il est évidemment passionnant et très utile de savoir sur quoi les gens travaillent, quelles sont les thématiques qui émergent, celles qui stagnent ou régressent, les contestations, mais aussi qui travaille et publie avec qui, etc. Dans cette approche, il faut se saisir des questions, et c'est essentiel, à un niveau agrégé (grands laboratoires, Universités, pays) et surtout pas au niveau individuel. En aucun cas utiliser ces indicateurs pour évaluer les individus, en particulier au moment du recrutement, et s'agissant de jeunes. On est loin des usages qu'on fait actuellement de la bibliométrie (annexe 9, VRS, 2008)

Alors que l'on peut raisonnablement estimer que les publications ne constituent pas l'alpha et l'oméga de l'activité de recherche⁴, le développement de la bibliométrie renvoie à l'incertitude radicale sur la nature du produit "recherche". Cette "incertitude s'alimente à deux sources, qui ne sont pas indépendantes : qui donc est demandeur de la recherche ordinaire, celle qui va se traduire par des publications qui disent d'elles-mêmes qu'elles sont "scientifiques" et comment juger de la quantité et de la qualité de la recherche produite par l'ensemble d'une Université ? Difficile de porter un jugement *d'ensemble*, qui requerrait une compétence exceptionnelle. On va donc descendre à un niveau plus micro, celui des chercheurs eux-mêmes, en écartant au passage les interactions entre chercheurs – si on ne les écartait pas, il faudrait porter un jugement *d'ensemble* sur la recherche produite au niveau (probablement) du laboratoire – et on serait renvoyé à l'impasse précédente, sous une forme à peine moins aiguë. Donc on finit par se résigner à cette forme... d'individualisme méthodologique extrême dans l'évaluation de la recherche : le comptage des publications, dans les revues dites "scientifiques" " (Annexe 6 Favereau, in Vatin, Caillé et Favereau, 2010).

1.4. Conformisme et danger pour la démocratie

Comme dans bien d'autres domaines, compter permet de ne pas débattre, et se révèle infiniment plus simple et plus rapide.

Même en admettant qu'il faille mesurer, compter, à trop vouloir accroître la productivité académique, le régime d'évaluation quantitatif affaiblit la productivité scientifique.

Ce régime d'évaluation "incite au conformisme des chercheurs, et génère une standardisation du travail de recherche conduisant à l'asphyxie progressive de la dynamique de la connaissance scientifique. On remarquera, qu'en économie, cette tendance est sensiblement renforcée par la dérive instrumentaliste de la discipline" (Annexe 3, Wallery, 2011).

Si elle accroît la productivité académique telle qu'elle est aujourd'hui mesurée, la sur-publication pèse négativement sur la productivité scientifique ; en outre, en favorisant l'inertie

⁴ "Les publications scientifiques ne sont pas le produit final de la recherche, mais un produit intermédiaire, certes essentiel, mais encore en attente des opérations, individuelles et collectives, qui le transformeront peut-être en un surcroît d'intelligence pour tous." (Annexe 6 Favereau in Vatin, Caillé et Favereau, 2010).

des spécialisations, elle réduit les capacités de renouvellement, et contribue ainsi à asphyxier ou stériliser la dynamique de la connaissance scientifique (Annexe 3, Wallery, 2011).

Parallèlement à un régime d'évaluation de la recherche qui incite au conformisme et stérilise largement la dynamique de recherche, peu d'attention est accordé à l'évaluation des enseignements dispensés à l'Université. Il est, de ce point de vue, remarquable d'observer que le devenir professionnel des étudiants (ceux qui échouent mais aussi ceux qui réussissent !) est assez mal connu et ne fait pas vraiment partie des critères d'évaluation des enseignants-chercheurs.

La généralisation d'évaluations chiffrées représente non un progrès, mais plutôt un danger pour la démocratie : d'une part parce que ces évaluations occultent tout ce qui n'est pas facilement mesurable ; d'autre part parce qu'elles risquent d'être utilisées pour clore des débats qu'il importe, en démocratie, de maintenir ouverts. Ce n'est pas ici l'opération de quantification qui est en cause mais bien l'autonomisation de ses produits (Annexe 5, M.-C. Bureau).

Parce qu'il est plongé dans l'histoire, le savoir, comme toute production sociale, ne saurait être réduit à une métrologie, si raffinée soit-elle. La métrologisation généralisée à dominante économique qui tend aujourd'hui à s'imposer n'est donc pas sans risques pour la cohésion sociale présente, mais aussi pour la capacité des sociétés à se penser dans le futur (Annexe 6 Vatin, Caillé et Favereau, 2010).

Le présent débat sur l'Université souffre tout à la fois de l'insuffisance gravissime de mesures et de réflexions métrologiques pour penser l'efficacité de nos systèmes d'enseignement et de recherche, et d'une illusion, tout aussi grave, sur l'idée que la question pourrait être réduite dans une métrologie économique universelle, qui est au final incapable de penser le savoir dans ce qu'il a de plus spécifique : la production du "nouveau" (Annexe 6 Vatin, Caillé et Favereau, 2010).

L'examen des sommaires et des *abstracts* (résumés des articles) des revues permet d'identifier relativement rapidement quels sont les sujets traités - à un moment donné - par les chercheurs (du monde entier ? on reviendra plus loin sur l'accès à ces titres et sur leur diffusion), et de dialoguer éventuellement avec eux, de les contacter (et plus si affinités...), bref de se repérer dans cet univers complexe qu'est notre "communauté scientifique". Cela permet en même temps de savoir quels sont les sujets "à la mode" - *i.e.* ceux qui ont été retenus par les revues patentées - ce qui peut être d'un grand secours pour postuler auprès des diverses "agences de moyens". Enfin, dans le même ordre d'idées, cela permet de se repérer dans le maquis des supports, pour identifier sans risques "ce qu'on fait de mieux" et tenter de s'y conformer pour y accéder ou y faire accéder nos doctorants.

(...) Pourtant, ce système recèle de graves défauts qui sont l'exact envers de ces avantages, mais qui vont bien au-delà. On voudrait ici insister sur un aspect insuffisamment évoqué, et qui ne se retrouve pas – ou moins – dans les sciences "exactes" : *le conformisme que ce type d'évaluation génère concernant les thématiques, les paradigmes, l'origine et le type de données mobilisées, les plans d'expositions et d'argumentation, et donc les façons de réfléchir, d'écrire et de convaincre, défauts qui annihilent largement les avantages potentiels de la forme "article"* (Annexe 10, Audier, 2009).

I.5. Une perte de pouvoir des instances de la profession

En déléguant la substance du pouvoir d'évaluation aux revues, le régime d'évaluation quantifié (REQ) leur confie, du même coup, le pouvoir de définir les normes et les modes propres à chaque discipline (Annexe 3, Wallery, 2011).

A été délégué aux directions des revues, à leurs *boards* (comités de rédaction) et à leurs *referees* (rapporteurs anonymes), ainsi qu'à leurs clients et lecteurs, le pouvoir de trier parmi les recherches – en déterminant celles qui sont dignes d'intérêt et celles qui ne le sont pas – et parmi les chercheurs – en distinguant les "excellents" des autres. Avec la bibliométrie, ce sont les supports de publication – en l'occurrence les revues, au détriment des autres modes de diffusion du savoir, tels les ouvrages – auxquels on transfère *ipso facto* le pouvoir exorbitant d'évaluer, l'AERES, l'ANR, et bientôt les dotations en crédits des Universités considérant la bibliométrie comme l'indicateur suprême de l'activité et de la qualité de la recherche (Annexe 10, Audier, 2009).

Les réformes imposées par le gouvernement ont semblé viser, *via* l'obligation de se soumettre à des évaluations quantitatives répétées, et sous prétexte d'autonomie des Universités, à déposséder définitivement la communauté savante de la part d'autonomie qu'elle conservait. Ce n'est donc pas l'existence d'une norme métrique qui fait en elle-même problème. Ce qui est problématique n'est donc pas l'introduction de la mesure en elle-même, mais deux choses assez étroitement liées : le fantasme de la bonne mesure unique (du *one best way*) et la liquidation de l'autonomie relative des corps professionnels qu'il favorise (Annexe 6 Caillé, in Vatin, Caillé et Favereau, 2010).

Enrayer la dérive quantitativiste du régime d'évaluation suppose, en premier lieu, que la substance du pouvoir d'évaluation ne soit plus accaparée par les revues mais reprise par les instances qui en sont officiellement dotées... Or, une telle ré-appropriation ne peut être envisagée que si les instances prévues à cet effet en reviennent à une évaluation sur travaux (et non sur indicateurs) (Annexe 3, Wallery, 2011).

I.6. Le pouvoir et la norme

Le traitement de l'incertitude par la mesure peut aboutir à une forme de captation de pouvoir cognitif, susceptible de générer des rentes économiques (Annexe 6 Vatin, Caillé et Favereau, 2010).

La dimension performative de ces instruments est primordiale, qu'elle soit considérée comme une simple explicitation des valeurs déjà partagées ou comme une assignation nouvelle.

On sait désormais que les classements peuvent modifier l'ensemble des pratiques de recherche et de publication (...). Le classement fonctionne comme une "technologie invisible", crée ainsi de nouvelles hiérarchies, et standardise les pratiques et les jugements (marginalisation de certaines revues, diminution de la diversité des méthodes et des objets) (Annexe 7, Pontille et Torny, 2010).

Les classements, et particulièrement ceux des "meilleures revues", sont utilisés pour mesurer la qualité des *departments* (facultés) ou des *business schools* (écoles de commerce) via le nombre de publications de leurs membres.

Ces listes ont été utilisées pour évaluer la production scientifique des Universités à partir de 1995 (Annexe 7, Pontille et Torny, 2010).

S'il est concevable que l'usage des listes de revues puisse servir les fractions dominantes des économistes et gestionnaires qui partagent une vision élitiste de leur discipline, nous estimons que l'effet dépréciatif est sans doute plutôt un produit inattendu d'un nouvel outil (Annexe 7, Pontille et Torny, 2010).

Suivant l'usage, il y a donc **les revues qu'on compte et les revues qui comptent**. Les revues que l'on compte permettent, dans le contexte actuel, d'établir un " périmètre" des revues acceptables par les instances d'évaluation. Les revues qui comptent produisent, au sein de ce périmètre, une "catégorisation hiérarchisée" (Annexe 7, Pontille et Torny, 2010).

L'effet de prescription du classement a été très important et très rapide chez les acteurs qu'il visait : soumission d'articles aux seules revues "qui comptent", ré-orientation des sujets de recherche, adoption des méthodes utilisées par les articles les plus "célestres", etc. La performativité de l'instrument est donc forte et se manifeste dans sa capacité à provoquer des alignements de jugement sur la qualité respective des revues (Annexe 7, Pontille et Torny, 2010).

Plusieurs limites du classement des revues en gestion comme en économie ont été épinglées. Ce classement a pu être jugé discriminatoire sur le plan linguistique puisqu'il valorisait principalement un petit nombre de revues de langue anglaise. La discrimination linguistique ne peut cependant être isolée du contenu même des recherches. Le risque d'une standardisation des travaux par l'alignement sur des critères internationaux -en fait nord-américains- a ainsi comme vecteur les revues "bien classées" de langue anglaise (Annexe 7, Pontille et Torny, 2010).

La démonstration a été apportée que les revues classées comme étant de rang 1 (1* pour les revues dites "généralistes"), toutes anglo-saxonnes, ne publient, en réalité, pratiquement que des universitaires appartenant à leurs cercles les plus proches, et que les contributeurs français en sont quasi totalement absents (sauf dans deux sous-disciplines spécifiques). Les français ont donc désigné comme cible d'excellence des revues... auxquelles ils n'accèdent pas, sauf réseaux particuliers construits avec des collègues américains ou assimilés ayant déjà pied dans ces revues en tant que membres d'un *board*, comme *referees*, ou comme *contributeurs* habituels (Annexe 10, Audier, 2009).

1.7. Le question-clé de la diversité

L'économie se trouve dans une situation critique face aux intérêts qui cherchent à l'instrumentaliser.

Selon Postel, Sobel et Tinel (2010)⁵, le risque qu'encourt actuellement le champ de l'économie est celui d'une rigidification des canons de scientificité autour d'un paradigme unique et celui, corollaire, d'une stigmatisation académique de tout autre positionnement théorique ou méthodologique. Il est aussi celui d'une fuite en avant de la théorie dans la technicité, afin de se conformer à l'image des sciences de la nature, au risque de couper l'économie des autres sciences sociales et de la priver des richesses de l'interdisciplinarité.

⁵ N. Postel, R. Sobel et B. Tinel, « Économie et pluralisme : l'Association française d'économie politique », éditorial, *Revue française de socio-économie*, 1^{er} semestre 2010.

Si l'on suit les trois auteurs, alors on peut avancer qu'enfermement idéologique et dérive instrumentale sont les deux maux qui guettent la discipline et sa production académique.

Il a toujours existé, dans la longue histoire des idées, des théories dominantes, des théories dominées, des théories émergentes, des théories progressivement abandonnées, une concurrence, au final, entre différents "paradigmes".

Le problème aujourd'hui n'est pas que le mainstream soit néoclassique ou libéral. Le problème se situe dans la domination sans partage de ce qui apparaît comme la "pensée unique" autorisée en économie. Le fait que de récents prix Nobel en économie (ou prix assimilés à un Nobel) aient été attribués à des penseurs ne relevant pas de ce courant (Sen, Stiglitz, Krugman, Ostrom) ne modifie pas ce constat. Celles qui sont posées comme les grandes revues de la discipline (classées 1*), les grandes Universités (notamment américaines), les principales institutions de la discipline sont aujourd'hui tenues par le courant néoclassique. A cela, deux explications sans doute. La première est l'appui dont ont bénéficié les représentants de ce courant depuis longtemps et singulièrement depuis la seconde guerre, ainsi que les moyens dont ils disposent, émanant de différentes institutions ou fondations proches des milieux d'affaires et de certains gouvernements. La seconde explication est le très grand émiettement de ce que l'on désigne parfois comme l'économie "hétérodoxe", constituée de multiples courants.

C'est donc aussi sous cet angle qu'il faut comprendre l'exigence du pluralisme. Cette exigence va au-delà de l'idéal de tolérance propre à toutes les activités intellectuelles. Le pluralisme est une manière pour la communauté des économistes de réfléchir à son rôle social et d'éviter la manipulation. Ce pluralisme, l'AFEP se propose de le décliner selon trois axes.

- Pluralisme des approches conceptuelles au sens où doit être acceptée l'idée que le débat sur ce que sont les concepts de base de l'économie n'est pas clos. À côté du courant néoclassique, existent d'autres traditions de pensée également fécondes, également légitimes, notamment l'institutionnalisme, le keynésianisme, le marxisme, l'évolutionnisme, etc. Cette diversité doit être considérée comme une richesse constitutive de l'économie.
- Pluralisme des points de vue, par quoi nous voulons signifier que l'économie ne se résume pas à la théorie au sens étroit. Il est d'autres manières de faire avancer les connaissances : l'histoire des faits et de la pensée économiques, la réflexion sur la méthodologie, la philosophie économique, la sociologie économique doivent être considérées comme des points de vue légitimes et importants.
- Pluralisme, enfin, des disciplines, car l'économie appartient aux sciences sociales et doit renforcer ses liens avec celles-ci. La situation actuelle qui tend vers une hégémonie extrême de l'économie à leur égard plutôt qu'une collaboration équilibrée est malsaine. La fermeture de l'économie trouve là une de ses racines les plus significatives. Les sciences sociales forment une entité structurée qui doit grandir de manière équilibrée dans ses diverses composantes pour insérer l'économie dans le réel, pour comprendre les faits économiques en tant que faits sociaux.

Une économie "polydoxique", c'est-à-dire admettant et encourageant la pluralité des approches, des méthodes et des objets de recherche, entre en écho avec la conception qui est

défendue ici de l'économiste au service du "demos"⁶ qui ne produit pas un savoir d'expert sous la forme d'une vérité prétendument dénuée de toute contingence (représentation implicite, intérêts défendus), mais qui contribue à éclairer les choix possibles et à en inventer de nouveaux.

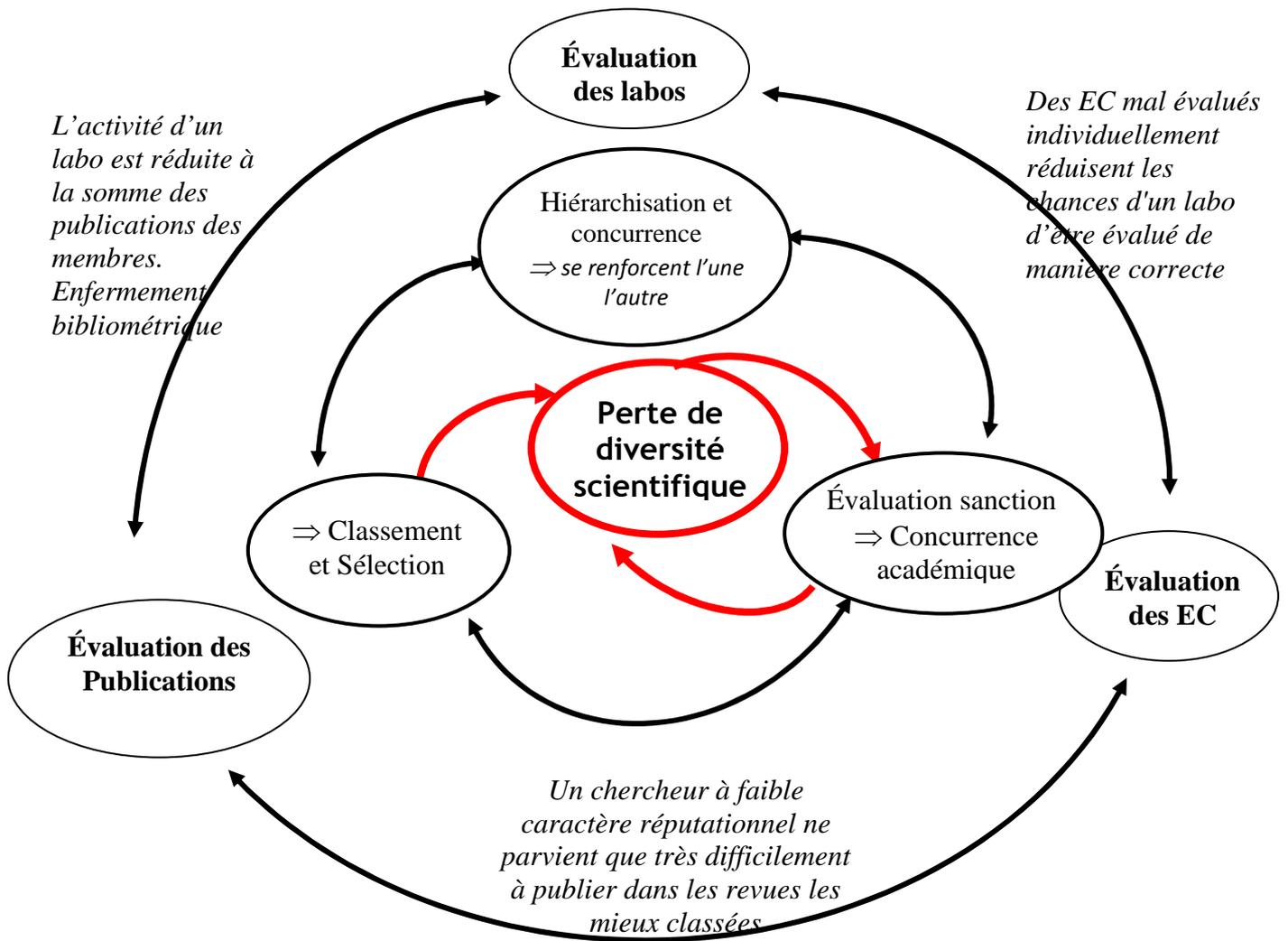
Le pluralisme est un enjeu clé pour les Sciences humaines et sociales (SHS), et donc pour l'économie, qui en fait partie. Il s'agit du seul moyen pour faire face à ce qui est souvent considéré comme l'échec des économistes face à la crise... voire leur contribution à l'engendrement de dérèglements, et l'incapacité de la discipline seule à proposer une lecture du monde.

Cet échec est d'autant plus marquant que l'économie passait jusqu'à la crise pour être la seule discipline au sein des sciences sociales à pouvoir prétendre au statut de science véritable.

⁶ Le peuple citoyen, celui qui débat et oriente les choix politiques

I.8. L'évaluation comme "système"

La congruence et même le caractère systémique du régime quantifié d'évaluation tel qu'il existe aujourd'hui et est basé sur les trois composantes que sont l'évaluation des laboratoires, l'évaluation des enseignants-chercheurs et celle des publications (classement hiérarchisé), peut être représenté comme suit :



Logique d'auto-renforcement :



Les mécanismes de hiérarchisation, de mise en concurrence sur les ressources et d'évaluation-sanction fonctionnent conjointement. Ils aboutissent à un rétrécissement du champ de la science économique et de la diversité scientifique.

Ce schéma illustre ce qui peut être considéré comme un référencement circulaire ou un processus en spirale d'autorenforcement aboutissant, au final, à une perte de diversité académique.

Cette perte de diversité a d'autant plus de chance de se produire que des conflits d'intérêt ou une certaine consanguinité sont probables entre évaluateurs et évalués : les enseignants-chercheurs qui sont referees A+ (les referees dont l'avis pèse le plus) des revues 1* (les revues qui comptent le plus) pèsent sur l'évaluation AERES. Cette évaluation agit -via le modèle SYMPA "Système de répartition des moyens à la performance et à l'activité"- sur la dotation des laboratoires des referees A+, etc.

I.9. Les effets du système d'évaluation sur chacune de ses composantes et leurs liens

I.9.1 Les enseignants-chercheurs

Tout pousse le chercheur à se focaliser sur le nombre de ses publications. C'est de cette obsession que résultent les pratiques de sur-publication (Annexe 3, Wallery, 2011).

Le chercheur "rationnel" devrait consacrer infiniment plus de temps et d'énergie à faire en sorte de publier, qu'à chercher.

Le REQ, conjugué à la systématisation des financements sur projets, rend ainsi le conformisme rationnel pour le chercheur (Annexe 3, Wallery, 2011).

La constitution d'un quasi *marché des chercheurs que les laboratoires tentent d'attirer ou de retenir à coup de primes, interventions diverses au moment des procédures d'avancement, décharges de cours*, repose sur une distinction essentielle entre :

- la désignation "basique" de chercheur produisant⁷ : un nombre de publications référencées (2 ou 4 selon que l'on est EC ou chercheur),
- surtout, la qualité reconnue des publications : ici, c'est la liste hiérarchisée de la section 37 qui compte.

La standardisation des comportements et du travail des chercheurs est évidemment nuisible à leur créativité, à leur imagination, à leurs ambitions, à leur esprit critique, à leur goût du risque...en incitant les chercheurs à consacrer leur énergie et leur intelligence à l'incorporation des normes et à la satisfaction des critères d'évaluation. Le conformisme qui en résulte génère d'importants gaspillages de ressources, du point de vue des progrès de la connaissance scientifique. Le développement de la connaissance scientifique, les capacités d'innovation des chercheurs et le renouvellement des disciplines, sont ainsi étroitement canalisés et bornés (Annexe 3, Wallery, 2011).

⁷ Comme ceci a déjà été observé, un chercheur publiant des articles référencés (si possible 1, voire 1*) est seul considéré comme produisant. La publication de livres, de chapitres d'ouvrages, la participation à une ANR ne suffisent pas –dans le régime d'évaluation actuel- à attester d'une production scientifique uniquement mesurée par la publication d'articles.

1.9.2 Les laboratoires

Le couplage entre l'évaluation des enseignants-chercheurs et celle des laboratoires est d'une évidence triviale si l'on considère qu'un laboratoire est une simple somme de chercheurs producteurs et non, centralement, une entité et un collectif.

Comme cela devient la règle dans la pratique (voir les cas grenoblois et toulousains, à titre d'exemple), l'évaluation d'un laboratoire - se fait en deux temps principaux, identiques à ceux de l'évaluation de l'enseignant-chercheur :

- la définition du périmètre des chercheurs producteurs ;
- la qualité des publications comptabilisées.

L'objectif résumé, caricatural, est de tendre vers 100 % de chercheurs producteurs, et le nombre maximum de publications de rang le plus élevé. Plus le laboratoire est proche de cet objectif, plus élevée sera sa dotation sur les crédits recherche.

"Il n'y a pas d'autre issue", déclare un directeur de laboratoire accablé mais contraint de "jouer le jeu".

Les nouvelles catégories d'évaluation ont rapidement été reprises par les évalués, déclenchant ce que certains ont appelé "la course aux étoiles". Estimant qu'ils n'étaient dorénavant plus jugés sur la lecture directe de leurs travaux, ni même sur le signal de qualité que constitue le nom d'une revue, certains laboratoires ont organisé leur rapport quadriennal autour du nombre d'étoiles (Annexe 7, Pontille et Torny, 2010).

1.9.3 Les liens entre l'évaluation des enseignants-chercheurs, celle des laboratoires et celle des Universités

Concernant à présent l'articulation entre l'évaluation des enseignants-chercheurs, celle de leur laboratoire et celle de leur Université, il convient de bien analyser les "demandes" adressées actuellement aux EC par leurs Universités respectives. Les textes sont clairs : l'évaluation des EC appartient aux sections du CNU, mais les financements du ministère aux Universités dépendent des résultats en recherche. Or ces résultats, dans le contexte actuel de rareté des ressources, entrent dans le calcul des dotations annuelles et donnent lieu à des négociations serrées entre les Présidents d'Universités et le ministère.

De plus, l'évaluation individuelle annoncée va se faire par vagues (quadriennale-quinquennale / mois de naissance). Si l'évaluation quadri/quinquennale l'emporte (ce qui est probable), certaines Universités ne disposeront donc des résultats pour leurs EC que dans plusieurs années (c'est le cas, par exemple, de Toulouse 1 et de l'UPMF de Grenoble 2). Il y a donc une sorte de "vide" que certains s'appêtent à combler en recourant à ce que nous pouvons appeler des "méta-règles" : les Présidents anticipent sur les résultats d'une évaluation dont ils ne disposent pas, en projetant le sens et les critères d'évaluation qu'ils connaissent. Ce "vide" est lié au fait que, dans le projet initial de décret concernant le CNU, il n'était pas prévu que l'évaluation individuelle soit faite par le CNU, mais par l'AERES. La pression des personnels et des organisations syndicales a eu pour effet de laisser le CNU en position d'évaluateur des EC (ce qui correspond à l'esprit de l'autonomie de la profession).

Sur un autre aspect du même point, les Présidents de telle ou telle Université (Grenoble 2, par exemple) qui retiennent le critère le plus strict de l'AERES (2 publications de rang 1 dans les

4 ans) ne sont ni en avance ni en retard sur les pratiques susceptibles d'être anticipées et mises en œuvre. En effet, si la cotation a disparu dans la liste AERES de juin 2010 et si la règle pour être déclaré "produisant" n'intègre pas la hiérarchie des revues, il reste à savoir si la liste AERES sera celle qui s'imposera pour les évaluations individuelles. Il est indispensable de préserver un périmètre ouvert et non hiérarchisé dans le cadre de l'évaluation individuelle, auquel seul correspond la liste AERES.

I.10. Au final, la nécessaire clarification de l'intention stratégique

La question de la qualification ou, plus exactement, de la clarification de **l'intention stratégique** de l'évaluation est essentielle, et apparaît comme la condition d'une démarche d'évaluation renouvelée.

Actuellement, l'évaluation qui domine est l'évaluation sanction fondée sur des principes ou méta-règles comme la concurrence, la sélection, le classement, l'auto-exclusion, la gestion de la pénurie (par les intéressés eux-mêmes).

Quand bien même accepterait-on certains de ces principes - celui de concurrence, par exemple - que des biais très importants ne peuvent être ignorés.

Comme cela a été évoqué, la rigidification des canons de scientificité autour d'un paradigme unique, la fuite en avant de la théorie vers la technicité produisent des biais dont il est difficile de penser qu'ils n'ont aucun rapport avec la production d'une forme de captation du pouvoir cognitif susceptible de générer des rentes, académiques comme économiques.

L'imbrication entre les modalités de l'évaluation, ses résultats et ses effets ne cesse de s'approfondir.

D'un côté, "fléchage" des promotions, des crédits, des postes, des financements et création de labels "d'excellence" ; de l'autre, soupçons de flânerie, stigmatisation des enseignants suspectés se réfugier dans l'enseignement ou dans l'écriture de livres, pire dans la participation à des débats publics à caractère idéologique... et sanction : menaces de modulation de service (sous la forme d'un alourdissement des tâches d'enseignement), enlisement dans le déroulé de carrière, diminution de la dotation recherche des laboratoires et menace de fermeture...

À l'opposé de ce mode d'évaluation qui ne sert qu'à dissimuler (mal) les intérêts qu'il sert tout en se prétendant neutre, objectif ou "scientifique", l'AFEP entend promouvoir une évaluation professionnelle fondée sur des principes alternatifs comme ceux de coopération, de promotion (des enseignants-chercheurs comme des laboratoires) de production de connaissances utiles.

Une telle évaluation, qui doit se baser sur la lecture et la connaissance du contenu des travaux évalués, est une évaluation "conseil". Il s'agit d'aider les enseignants-chercheurs et les collectifs scientifiques qui le demandent dans une perspective positive. L'évaluation n'est pas subie mais est attendue, elle n'est pas exogène mais participative.

Des relations de confiance entre évaluateurs et évalués se nouent ou s'expriment au moment de l'évaluation, puisque l'intention de l'évaluation n'est ni le jugement ni la régulation des moyens, mais la promotion de l'enseignant-chercheur ou du collectif qui demande l'évaluation. Une telle évaluation –aux antipodes de la plupart des exercices effectifs d'évaluation actuels nettement orientés par le souci de vérifier la conformité à l'égard d'une norme dominante- peut, le cas échéant, conduire à ne pas promouvoir tel enseignant-chercheur qui négligerait son travail ou une part de celui-ci (par exemple une participation adéquate aux activités pédagogiques)

Un aspect important et clivant entre les deux intentions de l'évaluation telles que nous venons de les mettre en évidence est la relation au temps.

L'évaluation sanction s'accompagne d'un rapport au temps marqué par la variabilité des règles, la multiplicité des exercices et des instances d'évaluation, leur caractère bureaucratique. Domine la perte de temps, conséquence inévitable de la multiplicité des évaluations et de l'instabilité des règles.

À l'opposé de l'évaluation sanction, l'évaluation professionnelle ou évaluation "conseil" mise sur le temps long de l'évaluation qui est celui de la lecture des travaux et des projets scientifiques. Les règles de l'évaluation sont stables et ne changent pas sans que évaluateurs et évalués ne le décident dans une perspective d'apprentissage et d'amélioration des méthodes de la recherche.

II. L'évaluation des laboratoires de recherche

Synthèse : Gabriel Colletis

Contributeurs : Florence Audier, Jean-Pierre Chanteau, Hervé Charmettant, Gabriel Colletis, Antonella Corsani, Ai-Thu Dang, Florence Jany-Catrice, Thomas Lamarche, Philippe Méhaut, Sandrine Michel, Martino Nieddu, Christian du Tertre, Serge Walery.

II.1. La réaffirmation du rôle central et de la diversité des missions des laboratoires

L'AFEP considère que les laboratoires doivent rester au cœur du dispositif de recherche dans la mesure où c'est à leur niveau que vivent les collectifs de travail – grâce à la proximité physique et aux interactions qu'elle permet – qui sont à la base de la production de connaissances scientifiques (et ceci, sans minimiser l'importance du travail en réseau).

Les réformes opérées ces dernières années renforcent les stratégies individuelles. La concurrence est perçue comme le meilleur mode de coordination des équipes, alors qu'au contraire la manière dont elle est organisée distend les coopérations scientifiques.

Il convient de réaffirmer avec force que la production des laboratoires ne peut être conçue comme la somme des "productions" individuelles.

L'agrégation des choix individuels pour construire une fonction de choix collective débouche, en effet, sur des apories insolubles (Annexe 6 Vatin, in Vatin, Caillé et Favereau, 2010).

Loin de privilégier – comme cela est actuellement le cas - des organismes tentaculaires constitués de rapprochements surfaits entre équipes et/ou entre individus, et qui ont souvent pour objectifs d'entrer comme "masse critique" dans le benchmarking de Shanghai, l'AFEP se prononce en faveur de véritables UMRs fondées sur la complémentarité des apports en connaissance autour d'un projet scientifique.

Le processus encouragé de congruence pour atteindre "l'excellence" conduit *de facto* à un oligopole de certains laboratoires (TSE, EEP, GREQAM...). Cette structure de marché est-elle efficace dans le temps ? Comment réagissent les "champions" de ces structures, souvent épris des bienfaits et des vertus de la concurrence, à cette contradiction ?

Plutôt que de rechercher des effets de taille qui n'ont jamais démontré leur utilité⁸, c'est un effet de densité et de diversité qui doit être privilégié. Un système trop concentré et dont la diversité se réduit est appelé à disparaître.

La vie des laboratoires de recherche en économie a été fortement affectée ces dernières années, et ce de multiples manières : pertes de solidarité au sein des unités en raison d'une pression accrue à la publication (rappelons qu'un laboratoire n'est pas une somme de chercheurs producteurs), incitations excessives aux regroupements souvent artificiels d'unités

⁸ Tout semble indiquer que la masse critique en matière de recherche n'est rien d'autre qu'une idée reçue, sans fondement empirique (annexe 1, de Grossetti M., Milard B., Eckert D., Bouba-Olga O. [2011]).

au nom de la lisibilité internationale⁹, multiplication des financements sur projet au détriment des nécessaires financements de base pour mener des recherches de long terme, replis disciplinaires favorisés par la LRU, etc.

Le travail réel, la production de savoirs utiles doivent retrouver toute leur place. Sans négliger l'apport de telle ou telle approche, théorie ou méthode, ce qui est important est bien l'utilité par rapport à la demande sociale, aux enjeux économiques contemporains. Observons cependant qu'un chercheur doit cependant se sentir libre de définir son programme de recherche en ne le conditionnant pas nécessairement à ce qu'il perçoit comme étant la demande sociale.

Doit également être validée l'hypothèse que cela ne peut se faire dans des cadres éthiques qui font parfois (souvent ?) défaut. Ce qui fait émerger la question suivante : dans quelle mesure les recherches menées sont-elles toujours acceptables éthiquement, en particulier dans les usages que l'on peut en escompter ? C'est évidemment une question délicate, mais si on parvient à ajouter un registre éthique aux fondements des jugements dans les pratiques professionnelles, un grand pas sera franchi, et il sera alors possible d'aborder enfin des questions épistémologiques dans les débats évaluatifs.

La mission des laboratoires de recherche est multiple et doit être réaffirmée.

Pour l'AFEP, ces missions se comprennent comme suit :

- Élaboration de questions de recherche ayant une finalité d'avancée des connaissances et/ou de production de savoirs utiles ;
- Production de dispositifs de réponse à ces questions mobilisant les différentes dimensions de la recherche : épistémologique, théorique, conceptuelle, méthodologique, empirique ;
- Formation des chercheurs par et pour la recherche (dont accueil des doctorants et post-doctorants) ;
- Fertilisation de l'enseignement par les méthodes et résultats de la recherche (dont articulation avec les masters) ;
- Diffusion des connaissances et du savoir vers des publics différenciés.

II.2. Clarifications quant à l'intention stratégique de l'évaluation des laboratoires

La première question qui est posée par toute démarche d'évaluation - au sens de l'établissement d'un jugement de valeur - est son "intention stratégique". Deux orientations s'opposent :

- Celle qui domine actuellement dans l'Université, mais aussi dans la plupart des organisations, en France comme ailleurs, consiste à imposer une logique de classement des équipes (voire des chercheurs) pour être à même de justifier leur sélection, notamment, vis-à-vis de l'octroi de financements (voire dans l'établissement d'une hiérarchie salariale et statutaire). D'où la généralisation de l'usage des notes (A+, A, B, C) et des expressions "équipe d'excellence", "laboratoire d'excellence". Dans cette conception de l'évaluation, domine la mesure qui conduit, au final, à la production d'une note qui, à l'image du prix dans la théorie néoclassique, est supposée résumer la totalité de l'information pertinente ;

⁹ Voir Annexe 1.

- Celle que l'AFEP retient consiste à concevoir et à mettre en œuvre un dispositif réflexif, dirigé vers l'activité de travail réel, afin de faire progresser la professionnalisation des équipes de recherche et des chercheurs, en identifiant les difficultés rencontrées, les avancées. L'évaluation par les pairs prévaut alors, dans une perspective consistant à aider le laboratoire à préciser son identité scientifique (spécificité et pertinence de ses apports à la recherche ; fiabilité de ses méthodes).

Autrement dit, la première conception est tournée vers l'établissement d'un classement venant justifier la concentration des moyens financiers et humain *via* les recrutements sur les laboratoires dits "d'excellence" ; la seconde, défendue par l'AFEP, vise la production de connaissances, en se centrant sur le débat scientifique et la dynamique d'apprentissage professionnelle des équipes au niveau de tous les laboratoires.

L'opposition entre "évaluation-sanction" et "évaluation professionnelle" (ou conseil) dans la finalité de l'évaluation se reproduit aussi par des choix au niveau des domaines évalués et des méthodes d'évaluation.

II.3. Les principaux principes de l'évaluation des laboratoires que l'AFEP défend

L'AFEP est favorable à l'évaluation des laboratoires ou des équipes, mais au sens défini précédemment – qui est d'ailleurs historiquement constitutif du mode de production des connaissances scientifiques, toutes disciplines confondues. A l'opposé des pratiques actuelles, l'évaluation doit être au service de la production des connaissances, et donc conduite de façon résolument réflexive et positive, permettant aux unités de mieux préciser leur politique et leur stratégie scientifique dans leur environnement.

Ce processus d'évaluation doit être rigoureux, pluraliste, démocratique et réellement contradictoire. Entre autres, la dimension démocratique, associant des représentants de la société aux scientifiques (même si ceux-ci doivent demeurer majoritaires), permet de discuter la pertinence des orientations de recherche. La dimension contradictoire – principe bien établi par les juristes – améliore la qualité de la conclusion de l'évaluation. Ceci suppose, notamment, que les unités évaluées puissent produire des réponses indépendantes (y compris de leur Présidence d'Université car aujourd'hui, le rôle de l'évaluation est d'autant plus important qu'elle sert aussi d'appui aux Présidences d'Universités pour la distribution des crédits de recherche, voire des postes).

L'AFEP est aussi particulièrement attachée à ce que les processus d'évaluation ne contribuent jamais à la formation ou au renforcement d'un "marché de chercheurs" tel qu'il émerge actuellement sur le modèle du marché des hauts dirigeants d'entreprise ou des stars du sport : le principe d'une notation des laboratoires centrée sur le nombre de publications de rang 1 ou 1* attire (par le jeu de primes et de décharges d'enseignement) des chercheurs qui contribuent à améliorer le "score" de publications du laboratoire. Plus le score est élevé, plus les financements qu'obtiennent les laboratoires concernés seront élevés... leur permettant d'attirer ou de conserver les chercheurs produisant 1 ou 1*. L'évaluation-sanction induit ainsi une sélection auto-renforçante, d'autant plus que les évaluateurs sont ensuite majoritairement recrutés parmi les laboratoires distingués.

L'AFEP s'oppose à une telle mise en concurrence entre laboratoires, et au développement de pratiques qui, sous couvert de "procédures qualité", sont improductives et injustes, aux

antipodes des principes de fonctionnement scientifique. Bref, l'évaluation ne doit pas servir à classer, sélectionner pour ainsi justifier la fermeture ou la désassociation d'un laboratoire.

La réflexion en matière d'évaluation doit porter à la fois sur le statut de l'autorité d'évaluation, sur les méthodes et sur les critères d'évaluation.

Une seule instance (l'AERES), constituée principalement de membres nommés et donc en rupture avec le principe d'évaluation par des pairs élus qui a prévalu jusqu'ici (du niveau national au niveau local) se voit attribuer cette compétence, de fait sinon de droit. Pourtant, rappelons-le, l'évaluation n'a de sens qu'en rapport avec des missions débattues démocratiquement et des moyens assignés correspondants à ces missions. D'ailleurs, les autres niveaux (ministère, EPST et Universités de rattachement) continuent à produire leur propre évaluation, spécifique ou calquée sur les critères de l'AERES – pas toujours bien compris d'ailleurs (cf. par exemple la confusion entre "produisant" et "publiant"). Les laboratoires et leurs chercheurs se retrouvent donc à devoir assumer la charge de rapports d'évaluation pour des institutions non coordonnées depuis la création non concertée de l'AERES¹⁰, et internaliser leurs contradictions puisqu'elles ne font pas respecter explicitement le pluralisme nécessaire des critères à évaluer (logique de complémentarité), mais ne s'alignent pas non plus sur une méthode commune (logique qui permettrait au moins de laisser le temps au chercheur de chercher autre chose que des moyens), même pas celle de l'Agence Nationale pour la Recherche qui se veut pourtant le référentiel. Le tout dans une opacité dommageable (voir par exemple la construction et l'utilisation du modèle SYMPA (Système de répartition des moyens à la performance et à l'activité) pour la répartition entre Universités des crédits nationaux de recherche ; etc.).

À ce jour, ce qui ressemble à un véritable "lobby" utilise un principe mono-dimensionnel pour mieux asseoir ses critères et son monopole. L'AFEP considère que l'évaluation d'un laboratoire doit être produite selon plusieurs dimensions, lesquelles ne sont pas toutes présentes au même titre dans les différentes unités (ce qui souligne la diversité des missions d'un laboratoire de recherche et pose la question des "métiers" de l'économiste).

Si une pondération de *certain*s critères (ceux qui sont mesurables) ne doit pas être exclue – pondération qui ne signifie donc en rien que l'on cède à la tentation du "tout quantification" -, celle-ci doit être équitable et explicitée. Surtout, elle ne doit pas servir à justifier une quelconque prépondérance des critères quantifiables sur l'évaluation par les pairs du contenu des travaux qui reste le mode que l'AFEP privilégie.

La production des critères d'évaluation – non quantifiables comme quantifiables - doit répondre à la fois aux missions auxquelles un laboratoire doit répondre (voir supra) et respecter ou garantir le pluralisme : pluralisme des approches, des méthodes et des objets.

Sans refuser l'existence de critères d'évaluation basés sur la production scientifique de chaque chercheur, l'AFEP considère qu'il convient de privilégier les critères attestant du caractère collectif de l'organisation et du fonctionnement de la recherche : séminaires communs, participation de plusieurs chercheurs à des projets de recherche financés, co-publications, etc.

¹⁰ Dans un contexte où le président de la République lui-même a ironisé sur les chercheurs qui ne sont jamais évalués (cf. vœux 2009 : <http://www.elysee.fr/president/les-actualites/discours/2009/discours-a-l-occasion-du-lancement-de-la-reflexion.6868.html>), il n'est pas inutile de rappeler que : un chercheur est déjà évalué régulièrement par son EPST ; ses textes pour publication sont déjà évalués par un ou plusieurs rapporteurs ; son laboratoire est évalué par le ministère et, le cas échéant, par un EPST...ce qui ne signifie pas cependant que tous les enseignants-chercheurs effectuent correctement leur double mission d'enseignement et de recherche !

Enfin, l'AFEP rappelle que parmi les principes de base de toute évaluation (mal respectés à ce jour), la transparence et, plus encore, la stabilité des critères d'évaluation sont deux principes essentiels. L'AFEP appelle donc à un affichage public de critères stables, désignés ex ante (avant leur mise en œuvre effective) et pouvant être débattus avec les équipes évaluées.

II.4. Les domaines et les méthodes de l'évaluation à finalité de classement des laboratoires (les pratiques actuelles)

Il convient sans doute de rappeler en introduction de ce point qu'en termes de résultat des évaluations, l'économie gestion est dans une position très spécifique. Le nouveau groupe "SHS 1" (regroupant, au sein des sciences sociales et humaines, les économistes et les gestionnaires considérés comme faisant de la recherche sur le double thème des marchés et organisation) combine l'un des plus faibles taux d'A+ et le plus fort taux de C.

Dans les pratiques actuelles, contrairement aux principes défendus par l'AFEP, l'évaluation se cantonne à l'appréhension des supports matériels des résultats de la recherche et à leur dénombrement, ce qui renvoie à une vision industrialiste de la recherche. Ainsi, dans des projets ANR, peut-on lire en lieu et place des tâches à décrire "tâche 1 = écrire deux articles pour des revues internationales très bien classées". La dimension *dénombrable* est ainsi considérée comme la seule tangible et supposée être "objective". Il s'agit de rendre compte du nombre d'articles dans des revues, elles-mêmes classées, du nombre de citations du chercheur puis de manière agrégée, au niveau du laboratoire, des chercheurs considérés comme producteurs.

Les dimensions non mesurables, non dénombrables des recherches sont écartées de l'évaluation. Dans cette logique, le contenu des recherches et de leur lien aux enjeux sociétaux, la pertinence des références théoriques fondant la problématique, l'adéquation de la méthodologie à la problématique, ainsi que la portée heuristique des résultats sont largement ignorés. Ces dimensions, demandant une investigation lourde et des compétences spécifiques de la part des évaluateurs, ne sont pas véritablement prises en compte.

Ce travail est renvoyé aux comités de lecture des revues académiques. Mais comme de leur côté, les revues sont classées en fonction de critères eux-mêmes uniquement quantitatifs (effets de réputation liés au nombre de citations) et de critères directement idéologiques ou politiques, aucune évaluation de la pertinence des recherches n'est engagée, seuls les critères quantitatifs apparaissant recevables.

Le panel de revues ne prenant pas en compte la diversité des options scientifiques et des enjeux économiques à traiter, une hiérarchie et une homogénéisation des revues sur la base de l'orthodoxie s'impose comme naturelle ou même scientifique. Les ouvrages ne faisant pas l'objet d'un passage devant un "comité de lecture" - comme les articles dans les ouvrages collectifs - sont rejetés comme ne relevant pas de l'espace du travail scientifique. Il en est de même des activités dites de valorisation auprès d'institutions, d'entreprises, de publics au sein de la société civile...

S'agissant des ressources des Universités françaises, Vinokur (2008) rappelle que le problème est celui de l'héritage d'un sous-financement. La LRU "répond" à ce problème en prévoyant, au niveau des laboratoires, le remplacement des dotations récurrentes par un financement par projets piloté par une agence de moyens (ANR) dont les membres sont nommés par... et non élus, ainsi qu'une évaluation de la production scientifique sur des critères (bibliométrie,

brevets...) pourtant assez largement disqualifiés dans la communauté scientifique internationale (Annexe 4, Vinokur, 2008).

Les effets pervers de cette conception de l'évaluation sont avérés, à savoir une normalisation de la recherche selon les idéologies dominantes et le mainstream scientifique :

- tendance à la marginalisation, voire à la disparition, des problématiques pluralistes et des équipes de recherche défendant l'inscription des sciences économiques au sein des sciences sociales ;
- affaiblissement des postures critiques sur les dynamiques économiques contemporaines.

Il convient de souligner une fois encore le caractère congruent et autorenforçant de toutes les évaluations. Le fait pour un laboratoire d'obtenir une ou des "ANR" (contrat de recherche financé par l'Agence Nationale de la Recherche), un "labex" (label supposé identifier et primer les laboratoires "d'excellence") ou autres "ex" (labels de "l'excellence" académique à défaut de preuve de qualité scientifique) entraîne de façon presque certaine une évaluation très favorable. Mais le fait d'avoir su obtenir une évaluation très favorable facilite plus que grandement l'obtention d'une "ANR", d'un "labex", etc. Tout ceci va de pair avec la capacité à placer plus facilement des articles, à se faire coopter dans des comités de rédaction de revues, à être invité, etc... Si un laboratoire parvient par un bout à entrer dans la pelote de l'excellence, les effets de renforcement se combinent et s'articulent avec le profit attendu de la participation à un quasi-marché des chercheurs. Sinon, c'est la spirale récessive menant à la fermeture de l'unité de recherche, durant laquelle les personnels se retrouvent dans une forte situation d'incertitude préjudiciable à leur santé et à leur production.

II.5. Constats concernant l'AERES et position de l'AFEP

De l'analyse statistique de l'évaluation des laboratoires par l'AERES (extraits de l'annexe 13), nous tirons sept enseignements qui peuvent être synthétisés comme suit :

Premier enseignement : une surpondération confirmée de l'activité de publication

L'AERES revendique l'importance de la rubrique "qualité scientifique et production" - de facto, le nombre et le rang des publications - dans la détermination de la "note de l'unité".

Deuxième enseignement : une corrélation curieuse entre le taux de producteurs et la "qualité de la production scientifique"

Le taux de producteurs n'est pas corrélé très fortement avec les notes des laboratoires, en particulier la note de "qualité scientifique et production" qui devrait avoir un lien fort avec cet indicateur mis en avant par l'AERES : le dernier laboratoire noté A+ a un taux de producteurs d'à peine plus de 60 %, alors qu'un laboratoire noté B a un taux de plus de 80 % !

Troisième enseignement : la variabilité de l'évaluation d'une vague à l'autre

L'économie-gestion avait été pointée du doigt comme étant la discipline où l'évaluation était la moins favorable ; il semble qu'une "adaptation" soit intervenue de la vague D (2010-2013) à la vague A (2011-2014).

Cette "adaptation" pose un problème d'équité puisque la "note" semble dépendre de la période pendant laquelle les laboratoires ont été évalués. À titre d'illustration, notons le glissement très remarquable de la note B vers la note A lors du passage de la vague D à la vague A. Plus de la moitié des laboratoires (52 %) ont ainsi la note B lors de la vague D. Ce pourcentage baisse considérablement pour la vague A, puisqu'il tombe à moins de 30 % (29 %). Corrélativement, alors que 15 % seulement des laboratoires obtiennent la note A pour la vague D, ce pourcentage monte à 50 % pour la vague A. Ces chiffres montrent une sensible évolution dans la répartition des notes, dont on ne sait pas si elle tient à des consignes explicites ou à des comportements plus ou moins spontanés des comités de visite.

Par contre, la réduction drastique de la part des A+ (une part divisée par 5 environ, de 19 % à 4 %) semble indiquer désormais l'aspect d'"excellence", voire d'exception, que l'AERES entend donner à cette note.

Quatrième enseignement : des éléments de ratios défavorables aux laboratoires en développement

Le taux de producteurs est réalisé en mettant au numérateur les "producteurs", au sens de l'AERES (soit A1), et au dénominateur la somme des "enseignants-chercheurs" et des "chercheurs" (soit N1 + N2 dans le tableau des effectifs). Il est à noter que ce rapport est déterminé sur une base instable. Tantôt, le rapport (numérateur comme dénominateur) est calculé en référence au bilan, tantôt il est calculé par rapport au projet. Il semble même que, parfois, est considéré au numérateur le nombre de producteurs du bilan et au dénominateur le nombre d'enseignants-chercheurs et de chercheurs du projet. Il s'agirait alors d'une logique de calcul pour le moins curieuse ne pouvant qu'aboutir à des écarts de résultat très néfastes pour les laboratoires en développement pour lesquels il est prévu que le nombre de chercheurs augmente tandis que le nombre de producteurs est, lui, borné par des effectifs faibles initiaux. Un exemple montre qu'un laboratoire, noté C, s'est vu reconnaître un taux de producteurs de 46 % alors que le calcul, à partir des chiffres du bilan, aurait donné un taux de producteurs de 67 %, ce qui aurait peut-être changé la note de l'unité...

Cinquième enseignement : une procédure d'évaluation entraînant des coûts rédhitoires pour les petits laboratoires

L'effet-taille semble recouvrir des phénomènes qui ne sont pas uniquement liés à la petite taille mais au fait que les laboratoires de petite taille sont pénalisés par les modalités actuelles d'évaluation et les critères utilisés. On peut cependant penser que ce type de laboratoire a sa place dans la recherche, en se positionnant sur des objets précis dans le cadre d'une évaluation moins univoque.

Sixième enseignement : des chercheurs au sein des laboratoires devant des dilemmes difficiles

L'engagement dans l'encadrement des jeunes chercheurs peut être pénalisant pour les résultats de l'évaluation des laboratoires telle qu'elle est pratiquée par l'AERES. Il peut paraître injuste que les chercheurs qui s'activent le plus pour permettre à l'Université de former des futurs chercheurs soient au bout du compte pénalisés en ne pouvant souscrire aux critères de publication qui sont aujourd'hui imposés.

Septième enseignement : une contestation à la fois large et profonde des évaluations des laboratoires par les laboratoires concernés

Cette contestation, à travers les droits de réponse des unités aux évaluations AERES, n'est pas marginale, car elle émane du quart des laboratoires de la vague A (6 sur 24) et de près de la moitié de ceux de la vague D (13 sur 27), dont nous rappelons qu'elle a donné des résultats nettement plus sévères.

La contestation n'émane pas seulement des laboratoires les plus mal notés, plus du tiers des laboratoires contestataires ayant été évalués A ou A+.

Les droits de réponse émanant des laboratoires peuvent être résumés ainsi :

- Une contestation de principe, en particulier le fait que l'activité de publication n'est qu'une mesure imparfaite de la production scientifique ;
- Le handicap de la pluridisciplinarité : les publications dans les revues non référencées par l'AERES ne sont pas prises en compte ;
- L'omission des chapitres d'ouvrages, d'articles publiés dans des revues étrangères non anglo-saxonnes, de l'édition d'ouvrages collectifs, de rapports officiels, des articles acceptés mais non encore publiés, les problèmes de date... ;
- La non prise en compte de facteurs explicatifs tels qu'un congé maternité, des responsabilités dévoreuses de temps, la proximité du recrutement...

Enseignement hors analyse statistique : une évaluation à géométrie variable

Il semble (à travers plusieurs exemples) qu'un certain nombre de dossiers de laboratoires ayant reçu une mauvaise note aient, au final, été "repêchés" hors du cadre "technique" officiel par l'AERES. Ceci montre – s'il en était besoin – le rôle essentiel du politique ou du rapport de forces. En cause donc, l'objectivité auto-proclamée, et en évidence, la confrontation de légitimités.

La géométrie des méthodes évaluatives semble elle-même très variable. À partir d'une même base objectivée, des critères différents peuvent être mis en avant dans l'évaluation des laboratoires.

Les arguments, en particulier chiffrés, mobilisés sont toujours ad hoc et en conséquence, malgré leur implicite (et illusoire) caractère objectivé, ils peuvent (ou devraient pouvoir) toujours être ouverts comme "espace de controverse".

Il est, bien sûr, difficile d'affronter un système chiffré, fonctionnant de manière ad hoc, car il est imprévisible et ajusté en permanence...

De l'ensemble de ces constats ou enseignements, l'AFEP tire les conclusions suivantes :

a) L'AFEP considère qu'une instance d'évaluation du type AERES – si elle devait être maintenue – devrait :

- répondre aux critères d'indépendance et de transparence ;
- être composée de pairs exclusivement élus ;
- être capable, en se centrant sur la mise en débat des méthodes d'évaluation, d'éviter – ou du moins ne pas alimenter – des processus pervers comme : la formation ou le renforcement de marchés de chercheurs (qui s'apparentent à des marchés de stars), le durcissement de la concurrence entre laboratoires, le développement de "procédures qualité" aux antipodes des principes de fonctionnement scientifique.

L'AERES est aujourd'hui traversée de logiques contradictoires.

L'AFEP constate que, dans le champ des sciences économiques et de gestion, l'AERES vient trop souvent en appui d'un lobby dont l'unique perspective est d'établir un monopole au nom de la "compétition scientifique".

L'AFEP considère que les conditions de constitution des comités de visite ne sont pas transparentes, et qu'elles n'offrent pas de garanties contradictoires. De plus, certains de ces comités n'ont pas une pratique conforme aux recommandations de l'AERES, notamment pour l'évaluation des "produisants".

La pratique de certains comités de visite indique qu'ils ne se sentent non pas en charge d'une évaluation, mais investis d'une mission (qui n'est pas énoncée) de cadrage conduisant à l'instauration d'une uniformité dans les politiques scientifiques, auxquelles les laboratoires sont de facto sommés de se conformer (recentrage thématique, voire monothématique imposé ; intégration forcée de laboratoires ; rapprochement de laboratoires de province avec les écoles de commerce ; rôle central des incitations financière ; remodelages d'équipes sur la base du job market ; etc.). Il n'est pas rare d'observer que la partie du rapport de certains comités de visite concernant le projet scientifique du laboratoire tient en un nombre limité de lignes, dont la teneur relève de plus de l'injonction.

b) L'AFEP récuse ce système malthusien, fondé sur la notation, dont l'objectif essentiel est de réserver la possibilité d'accéder aux notes les plus élevées à un nombre très réduit de laboratoires. Elle souligne les distorsions dans la "sévérité" de l'évaluation que cela induit et qui, fréquemment, pénalise les laboratoires d'économie et gestion dans les Universités pluridisciplinaires. Elle souligne les inégalités entre les "vagues" quadriennales, les critères évoluant au cours du temps. (Voir en annexe 13).

c) L'AFEP constate que l'AERES s'appuie sur la liste CNRS de revues et son classement établi dans un autre objectif (discuter de l'avancement des chercheurs CNRS aux fonctions de directeur de recherches) pour invalider les stratégies de recherche et de publications de certains laboratoires. Notamment, il advient que ne soient prises en compte que les revues de catégorie 1 et 2 du CNRS pour ce qui concerne l'évaluation de la "qualité" des laboratoires. Ainsi il est arrivé que des revues soient explicitement désignées comme non pertinentes ou de niveau médiocre, alors qu'elles sont dans le champ même dans lequel s'inscrit la recherche¹¹.

¹¹ Inversement, à quelle logique peut bien répondre l'usage d'une catégorie "A+" alors que la catégorie A a déjà été construite comme celle de l'excellence (revues internationalement "reconnues" par la communauté scientifique) ?

En résumé, l'AFEP est hostile à une AERES de type "agence de notation" et s'interroge sur la capacité de cette agence à se réformer dans un sens qui serait de nature à promouvoir le pluralisme dans toutes ses dimensions ainsi qu'une recherche au service de la société et non d'intérêts catégoriels.

II.6. Préconisations générales

En cohérence avec sa conception de l'évaluation, et à l'instar du CoNRS¹², l'AFEP entend que soit mis fin à la "notation" qui ne peut que déboucher sur une distribution de "bons" et "mauvais" points, donc sur un classement et, au final, des sanctions (fermeture de laboratoires, désassociations, baisse de la dotation).

L'AFEP dénonce le fait que les jugements ainsi prononcés et rendus publics sont sans appel. Elle rappelle les conséquences de ces jugements : ils cristallisent des situations pour 4 ou 5 ans, ce qui – outre le caractère injuste et les effets de renommé qui y sont attachés – est contraire à une démarche d'évaluation positive.

L'AFEP estime que toute évaluation d'un laboratoire de recherche doit partir de l'énoncé de ses missions (déjà explicité dans le premier point de cette partie) :

- Élaboration de questions de recherche ayant une finalité d'avancée des connaissances et/ou de production de savoirs utiles ;
- Production de dispositifs de réponse à ces questions mobilisant les différentes dimensions de la recherche : épistémologique, théorique, conceptuelle, méthodologique, empirique ;
- Formation des chercheurs par et pour la recherche (dont accueil des doctorants et post-doctorants) ;
- Fertilisation de l'enseignement par les méthodes et résultats de la recherche (dont articulation avec les masters) ;
- Diffusion des connaissances et du savoir vers des publics différenciés.

L'AFEP propose que la conduite des évaluations :

- Reflète la pluralité des missions des laboratoires ;
- Ait pour finalité d'aider à l'affirmation de leur projet scientifique ;
- Garantisse, par le caractère démocratique de leur élaboration et de leur utilisation, le pluralisme des approches, des objets de recherche, des méthodes ;
- Exprime de façon privilégiée la dimension collective de l'organisation et du fonctionnement de la recherche ;
- Favorise l'ouverture disciplinaire.

L'évaluation pluraliste des laboratoires proposée par l'AFEP recouvre les items suivants :

1. La prise en compte de la dimension longue du temps de la recherche
2. L'évaluation par les pairs
3. Le principe essentiel de lecture des travaux
4. L'usage à modérer fortement des indicateurs bibliométriques

¹² Comité national de la recherche scientifique (CoNRS). Placé auprès du CNRS, le CoNRS regroupe des instances composées d'experts nationaux et internationaux ayant pour mission l'évaluation des chercheurs, l'analyse de la prospective des unités dans le cadre de la stratégie du CNRS et le conseil aux directions du CNRS.

5. Le soutien de thématiques de recherche innovantes
6. La prise en compte des apports méthodologiques
7. Un soutien des laboratoires en difficultés ou en démarrage

Ces sept items sont explicités et détaillés dans la partie (IV) du présent document consacrée aux préconisations.

II.7. Préconisations concernant l'AERES dans l'éventualité de son maintien

L'AFEP considère que l'évaluation des UMR et des autres laboratoires doit être conduite en fonction de leurs programmes et projets scientifiques, qui comportent selon leur histoire et leur implantation plusieurs dimensions. De là découle le refus de déléguer l'évaluation à d'autres "communautés", qu'il s'agisse des comités de rédaction des revues (qui sont étrangères aux objectifs, programmes et contraintes des équipes) comme à des comités de visite qui seraient chargés d'appliquer une "doxa" et de décerner des notes et classements.

C'est sur le maintien ou le rôle futur de l'AERES que les analyses des membres de l'AFEP auront été le plus diversifiées. Certains membres de la Commission s'accordent à penser qu'une réforme de l'AERES est impossible, voire que son existence n'a pas de légitimité, l'évaluation devant soit être décentralisée (comités choisis, composés par les laboratoires évalués en fonction de leur projet scientifique), soit revenir aux instances nationales (Comité national du CNRS et CNU). D'autres estiment une réforme de l'AERES possible à condition que soient repensés la composition, les missions, les critères d'évaluation de l'agence qui devrait elle-même faire l'objet d'évaluations régulières par un comité ad hoc composé d'élus de la profession. Au delà de cette différence d'appréciation, tous les membres de la Commission ont cependant convergé pour dénoncer l'existence d'une institution fonctionnant comme une agence de notation.

Dans la perspective d'une très sensible transformation des principes et procédures d'évaluation (résumée dans le point 6), et dans l'hypothèse de l'éventualité du maintien d'une institution de type AERES profondément rénovée dans cette perspective, l'AFEP entend que soit instaurée d'ores et déjà et systématiquement une phase de restitution des visites, sous forme de débats entre les membres des comités de visite et les membres des laboratoires (ou au moins des membres des conseils de laboratoires). L'envoi d'une note ou d'un simple rapport sans discussion sur le futur constituent au mieux une caricature de ce que devrait être une évaluation "positive". Ces pratiques devront donc être prohibées.

L'AFEP propose que la constitution des comités de visite relève d'une communauté reconnue et que tout laboratoire ait un droit de récusation. Si maintenus, les comités AERES devraient être composés d'élus et non de nommés. La présence d'élus est une garantie du pluralisme dans la mesure où ce qui sort des élections est, sauf situation particulière, un gage de représentativité du milieu. En aucun cas il ne devrait s'agir d'experts ou de « professionnels de l'évaluation », qu'il s'agisse des français comme des étrangers.

L'AFEP demande qu'une instance composée d'élus examine et rende compte régulièrement des pratiques de l'AERES dans le domaine des sciences économiques (et de gestion). Le principe est très ancien : les évaluateurs doivent être eux-mêmes évalués.

III. L'évaluation des EC

Synthèse : Pierre Alary, Thomas Lamarche

Contributions de : Pierre Alary, Florence Audier, Gabriel Colletis, Ai-Thu Dang, Florence Jany-Catrice, Thomas Lamarche, Sandrine Michel.

Deux grands axes de redéfinition des obligations professionnelles des chercheurs et enseignants-chercheurs (EC), opérant sur une assez longue période, en lien étroit avec la massification de l'enseignement supérieur, ont abouti à une transformation du métier : la division du travail et la spécialisation des enseignants. Des rapports successifs (Espéret 2001, Belloc 2003, Goulard 2007, Schwartz 2008) ont défendu la spécialisation au sein des enseignants-chercheurs, notamment en fonction d'une évaluation des publiants et des non-publiants, du volume d'heures d'enseignement et des responsabilités collectives et pédagogiques. Ceci constitue une voie de la fragmentation du statut et de la transformation de la profession qui ne prévoit que deux missions : l'enseignement *et* la recherche. Des changements récents, dans la lignée de la LRU et de la création de l'AERES, ont abouti à une codification nouvelle associée à une quantification de l'activité qui surdétermine une des missions des EC, celle de recherche, entendue selon des critères discutables et discutés. Et cela, aux dépens de l'enseignement (et des responsabilités collectives).

Cependant, l'ensemble des activités de recherche n'est pas pris en compte dans ce processus d'évaluation. L'activité de recherche est réduite à l'activité de publication, elle-même quantifiée par les articles publiés dans une liste de revues hiérarchisée. Ce mode d'évaluation des EC pose ainsi de nombreux problèmes. D'une part, cela ne permet pas à la profession d'assumer conjointement les missions de l'EC, en conduisant à la spécialisation et non à la pluralité des missions assurées. D'autre part, la diversité des méthodes et des thèmes de recherche, et plus largement la prise en charge de la dimension collective de l'activité ne sont pas correctement prises en compte. Privilégier la recherche, puis un type de recherche et enfin une catégorie étroite de publication présente deux inconvénients majeurs pour la profession :

- Premièrement, l'enseignement et les tâches collectives (administratives et d'animation de l'enseignement et de la recherche) ne sont pas véritablement reconnus, notamment pour certaines étapes significatives de la carrière (notamment pour le passage de maître de conférences à Professeur des universités). Consacrer de l'énergie à ces missions revient à "pénaliser" certains EC, alors que d'autres, en quête de "bonnes" évaluations, s'en écartent. Ces dispositifs d'évaluation introduisent une concurrence entre EC et entre les missions de l'EC. Cette concurrence nourrit une spécialisation qui remet en cause la pluralité des missions et qui risque de conduire à la mise en place de deux modèles de carrières.
- Deuxièmement, le critère dit "d'excellence" oriente vers une méthodologie singulière et vers un choix de paradigmes. La définition de l'excellence produit une sélection au profit d'une élite étroite, et élimine progressivement le pluralisme et le débat contradictoire, qui sont nécessaires à une dynamique de recherche et d'enseignement à même de faire face à des questions de société complexes (crise, chômage, pauvreté, politiques publiques...). Le pluralisme dans une discipline en étroit lien avec la société est une garantie de démocratie ; la stratégie dite "d'excellence" lui est adverse.

Le classement des Universités – encore nommé "palmarès" – pourrait être un instrument supplémentaire de fragmentation des missions des EC et des Universités. En effet, le ministère, se basant sur une enquête (dont la méthodologie est contestable et contestée), a procédé à un classement des Universités selon leur "performance" en matière d'insertion professionnelle des diplômés sortant de master. Le 1^{er} palmarès des Universités (publié dans le Figaro Magazine) procède à un classement par discipline, le critère retenu ici étant l'emploi à 30 mois, érigé en critère d'excellence quant à la performance de la professionnalisation, versus l'excellence en recherche, avec l'idée sous-jacente de déboucher sur une double hiérarchie : les Universités de recherche (sur le modèle des "research universities") et les Universités orientées "marché du travail" (sur celui des "college universities"). Les EC travaillant dans les unes ou les autres pourraient bien être régis par des règles différentes...

III.1. Le primat de la recherche sur les autres missions

Les sciences économiques ont abandonné l'idée d'évaluer l'activité même de recherche, comme activité complexe, lente, longue, et s'en tiennent non pas à l'évaluation des résultats, qu'il faudrait évaluer, ni même à l'ensemble des publications qui sont très diverses, mais à une forme de mesure à partir d'une liste de revues. La première position de l'AFEP est le recours à une évaluation de ce que font les EC, c'est à dire de leur activité. Pour cela, la première étape est la prise en compte de l'ensemble de publications de façon générale, et plus spécifiquement de l'ensemble des productions scientifiques.

Dans le fond, l'AFEP considère nécessaire de revenir à une évaluation qui parte de la prise en compte des travaux effectifs et de leur lecture par les pairs évaluateurs.

III.1.1 *Le rôle décisif de deux niveaux de listes qui trient et hiérarchisent*

Les dispositifs d'évaluation qui s'imposent, en concurrence, sont structurants. Ainsi la liste des revues de l'AERES est un instrument destiné au comptage des "produisants". Les listes, plus courtes, portées par les écoles de commerce et les fondations (EEP, TSE...) ont pour fonction d'allouer des moyens et des primes. Enfin, la liste de la section 37 du CNRS a tendance à être utilisée comme un outil de gestion des carrières selon un modèle professionnel (Pontille and Tourny 2010). L'évaluation des activités de recherche portée par l'AERES actuelle repose sur ces deux éléments : une liste qui définit le périmètre des chercheurs, et une liste qui hiérarchise les unités.

La liste AERES sert à identifier les EC "produisants" et "non producteurs". La sélection des producteurs est claire : tout EC ayant publié deux articles, au cours des quatre années du contrat dans des revues référencées dans la liste AERES, est considéré comme producteur. La liste est relativement large et ce type d'évaluation ne constitue plus le principal objet du débat, même si l'usage d'une liste mono disciplinaire pose problème, et même si la méthode revient à nier l'activité d'évaluation des travaux eux-mêmes, ce qui impliquerait au moins de les lire. En revanche, elle ne prend pas en compte les travaux publiés dans les ouvrages collectifs, les livres et les manuels. Par exemple, un enseignant chercheur ayant écrit un manuel destiné aux étudiants et un ouvrage de vulgarisation est *stricto sensu* et arithmétiquement considéré comme non producteur. Il y a là une vision tronquée de l'activité de l'EC, particulièrement pour une science des SHS qui a pris sur les débats démocratiques.

- Cela produit une coupure entre les missions d'enseignement et de recherche, alors que le lien enseignement / recherche est central dans l'enseignement supérieur de façon à ce que la dynamique d'enseignement soit en prise avec le savoir et les débats académiques.
- Cela ouvre potentiellement vers une politique de déclassement des EC non "produisants" dans une catégorie d'enseignants dont le service pourra être alourdi.
- Nous défendons la pluralité des parcours, des temporalités de la recherche et des manières d'associer les différentes missions, notamment en développant une activité de recherche en lien avec la pédagogie, et articulée aux débats de société (publication d'ouvrages académiques, de manuels, de livres, de rapports...)...
 - Prise en compte des temporalités de la carrière et du caractère irrégulier des productions et résultats scientifiques ;
 - Prise en compte de la parité, c'est-à-dire des effets déstructurants qui affectent la carrière des femmes ;
 - Possibilité donnée aux jeunes enseignants-chercheurs et aux enseignants-chercheurs expérimentés qui le souhaitent et qui en font la demande d'obtenir un allègement transitoire de leur service d'enseignement afin de leur permettre de franchir une étape en matière de recherche, étape qui sera évaluée.

La liste des revues de la section 37. Cette liste ne permet pas officiellement d'évaluer les EC. En revanche elle constitue implicitement la référence dominante, si ce n'est exclusive, en matière de qualité des travaux. Dans ce contexte, l'évolution professionnelle des EC dépend des critères qualitatifs qui structurent cette liste. Un doctorant "excellent", un docteur "excellent" sont censés publier dans des revues 1*, 1, voire 2. Or les revues dites "excellentes", la plupart anglo-saxonnes et porteuses de leur propre histoire, publient des travaux dont les fondements théoriques et méthodologiques sont orientés. Les revues publiant des travaux singuliers, dont les méthodes et les paradigmes adoptés s'éloignent de ces critères dits d'excellence, sont mal classées. Ainsi, produire dans ces revues (par exemple classées 3) peut bloquer la carrière des EC ayant fait de tels choix méthodologiques. Cela affecte les doctorants ; les docteurs au moment de la qualification et du recrutement ; les candidats, notamment MCF et CR-CNRS lors des passages à des postes de PR ou de DR. Tous sont évalués à l'aune de cette hiérarchie de revues. Dans la mesure où les revues les mieux classées appartiennent à une composante du champ disciplinaire (les revues les moins bien classées aux autres composantes de ce champ), la promotion passe par l'adhésion au premier champ cité. Cela pousse à un renforcement disciplinaire et disqualifie plus ou moins directement ceux qui s'en écartent. La liste est ainsi érigée en instrument d'un monopole académique.

Cette liste est instrumentalisée, et utilisée par certains pour faire prévaloir une politique d'écramage qui prend le masque de l'excellence. Ce qui est antinomique avec la nécessité de faire vivre une communauté diverse qui intègre les aléas de la recherche et les rythmes de la production scientifique.

La liste AERES a pour objet de définir qui est dans le périmètre, et de facto la liste CNRS est utilisée pour définir qui progresse (i.e. carrière) et en même temps qui a accès aux moyens. En effet, le classement des unités de recherche a une incidence directe sur l'accès aux ressources universitaires, aux contrats de recherche, aux financements ANR, mais aussi sur l'habilitation des écoles doctorales et des masters.

Il est nécessaire d'évaluer l'activité de recherche en lien étroit avec les moyens disponibles et l'environnement dans lequel elle se développe. Traiter de performances brutes, en dehors des ressources et des contraintes réelles, revient à nier les contextes et les conditions de l'activité.

Les différences entre Universités sont ainsi un élément indépassable de l'analyse de *l'excellence* et de la structuration implicite des EC en différents profils ou spécialisation.

III.1.2 Resserrement autour d'une élite éloignée de la communauté

Domine actuellement une vision malthusienne qui conduit à produire une élite hyper restreinte. Le déclassement potentiel de nombre d'EC, le repli vers une mission unique d'enseignement (mission *de facto* dévalorisée) et de fait le resserrement des thèmes de recherche sont le revers (...ou le but ?) d'une stratégie d' "excellence" fondée sur la seule activité de publication dans des revues classées 1 ou 1*.

Sur approximativement 2 500 chercheurs en économie (enseignants-chercheurs et chercheurs), seuls 5 % à 10 % sont "internationaux" : quelques "stars" seulement sont visibles ("excellents"), compte tenu du critère de publication. Cette norme ne profite qu'à un très petit nombre, qui réussit à la faire prévaloir pour tous. Le "top 5" des revues selon les classements en cours n'est pas accessible à la très grande majorité des enseignants chercheurs.

- Évaluation des revues : sur les revues d'élite (57 titres), depuis 2000, seulement 24 articles publiés par des Français... (voir annexe 10)
- Indices Combes-Linnemer : une stratégie de hiérarchisation.

Ce type de démarche vise à offrir une image recomposée du périmètre de la discipline dans une conception resserrée qui ne permet pas de faire vivre la pluralité des thèmes, des méthodes et des objets. Les indices partent du *Journal of Economic Literature* (JEL) et de ses 1 200 revues dont 1 064 actives, des 401 revues de la liste CNRS et des 680 revues de la liste AERES (liste antérieure à celle de juin 2010). La liste JEL est renseignée par différents indices de citations (Journal Citation Reports (JCR) 304 revues en sciences sociales dont 159 classées en économie (aucune revue française) + Google scholar).

Ces indices aboutissent à une hiérarchie en 3 niveaux :

- o 1er cercle = les publiants dans les 159 revues EconLit JCR Eco
- o 2ème cercle = les publiants dans les 304 revues EconLit + JCR
- o 3ème cercle = les publiants dans les 1 064 revues EconLit actives

L'intersection entre les quatre bases de revues en SHS (AERES, ERIH (Europe), SCOPUS, WOS-ssci) n'est que de 9,6 %, c'est-à-dire que seuls 9,6 % des titres figurent dans les 4 bases, d'où un problème majeur de représentativité des travaux publiés dans les revues consignées dans les unes et les autres.

"Aussi, étudier les classements de revues dans ces disciplines [économie et gestion] présente un troisième intérêt : les acteurs concernés mobilisent des savoirs, sinon plus spécialisés, au moins segmentés d'une manière beaucoup plus marquée que dans la plupart des autres SHS. (...) Plus que pour d'autres SHS, les négociations autour des classements de revues mettent en jeu la définition et les limites des groupes professionnels." (Torny et Pontille, 2010)

Être "excellent" selon ce mode d'évaluation ne dépend pas nécessairement du "mérite" des EC. L'excellence sélectionne avant tout sur un type de profil, très formaté, ayant de plus opéré des choix méthodologiques et paradigmatiques relativement précis. Le mérite ne s'exprime qu'à l'intérieur de ce cadre. Ce mode d'évaluation a donc un impact significatif sur la carrière

des EC dont la promotion dépend de l'adhésion au format. Ce dernier sert implicitement d'aiguillon aux doctorants, aux docteurs à la recherche d'un poste et aux MCF en exercice.

Le cas de la Prime d'excellence scientifique au CNRS¹³

La Prime d'excellence scientifique (PES), qui existe pour les chercheurs CNRS depuis 2010, cherche à promouvoir le principe de l'individualisation des carrières. En 2009, tous les CoNRS des différentes sections en SHS ont refusé de participer à la sélection des dossiers pour l'attribution des PES. Quant à la section 37 (Économie et gestion), si elle a refusé, comme beaucoup d'autres sections du Comité national, de participer à la sélection des dossiers, elle a néanmoins décidé de proposer aux chercheurs entrants de candidater. Le compte-rendu de la session d'automne résume ainsi sa position : "un jury ad hoc par institut a distribué les PES en proportion des demandes. Un membre de la section a été envoyé en tant qu'observateur. La section n'a effectué aucun classement, mais a proposé que soient prioritaires les derniers CR recrutés, auxquels elle a conseillé de candidater. Au niveau du CNRS, le taux de participation a été inférieur à 20 %. Il y a eu 40 candidats en section 37 sur à peu près 200 demandes en SHS¹⁴. Pour le moment il y a maintien du statut quo, aucune procédure pérenne n'a été mise en place". On constate en 2011 une baisse significative de dossiers de candidature dans toutes les sections, y compris dans la section 37. Dans la section 37, on a 24 demandes pour 9 PES. La chute du nombre de candidatures dans toutes les sections montre que les chercheurs ne semblent ni adhérer à la PES (qui est censée compenser des problèmes de salaires) ni, a fortiori, à ses conditions d'attribution. Conditions qui impliquent pour les chercheurs d'effectuer un service d'enseignement et font l'objet d'un contingentement.

III.2. La carrière : étranglement autour d'une seule mission, évaluée de façon uniforme

L'évaluation des EC pose le double problème de l'évaluation de l'ensemble des missions et de l'évaluation spécifique de la recherche (mission évaluée de façon discriminante). Pour ne pas être mal évalués, les EC sont conduits à délaisser leurs activités d'enseignement et les responsabilités collectives, pour se concentrer sur la publication de travaux formatés.

III.2.1 La thèse

Les doctorants mettent en œuvre des stratégies dont l'objectif est double. D'une part, ils cherchent à réaliser une thèse en trois ans comme dans les autres disciplines. Dans ce but, et pour majorer leurs chances d'intégrer l'enseignement supérieur ou la recherche, les doctorants sont poussés à standardiser leurs travaux selon le schéma connu : une introduction, une conclusion et deux à trois papiers. D'autre part, ils pensent leurs papiers comme des articles qu'ils commencent à publier avant la fin de thèse, souvent en co-signature avec des "seniors", pour avoir plusieurs

¹³ La rédaction d'un paragraphe consacré à la PES au CNRS ne signifie pas – tant s'en faut – que les conditions d'attribution de la PES dans les Universités seraient plus transparentes ou plus équitables.

¹⁴ Les demandes ont donc fortement excédé les entrants, mais les PES n'ont pas été attribués aux plus "gradés", contrairement à ce qui s'est fait dans plusieurs autres sections des SHS qui ont participé à la procédure d'attribution.

publications en fin de thèse. La visibilité du travail des doctorants repose donc dorénavant pour une grande partie sur le rang des revues dans lesquelles ils auront publié. Au niveau de la thèse, la norme implicite d'évaluation présentée supra joue également un rôle déterminant.

Cela induit un changement des critères d'accès à la profession, accompagnant le fait que les thèses changent de nature : d'une rédaction d'un travail de recherche conséquent et original, qui s'intègre dans un contexte référencé et analysé, on passe dans la plupart de nos sous-disciplines à l'agrafage de quelques (trois le plus souvent) articles souvent co-signés (parfois avec le directeur de thèse), le tout tenant lieu de thèse.

La thèse reste bien sûr la condition d'accès aux fonctions d'EC et de chercheur. Les thèses institutionnalistes risquent d'être progressivement battues en brèche non sur le terrain intellectuel, mais par tarissement du nombre de directeurs de thèse, et par la faible activité d'encadrement doctoral d'une génération d'encadrants potentiels. À noter que la faiblesse du nombre de directeurs de thèse est directement lié au rétrécissement des possibilités d'accès au "rang A", celui de Professeur des universités ou de Directeur de recherches au CNRS.

Les directeurs de thèse

Au cours de la mandature de quatre ans (2008-2011), on compte 103 directeurs de thèse "institutionnalistes" sur environ 577 (estimation les 2 dernières années).

Résultats agrégés (Nb de directeurs institutionnalistes = 103)	Nb	%
Dossier n'allant pas à la qualification (dossier non évalué)	12	12
Directeur sans candidat qualifié	42	41
Directeur ayant au moins 1 qualifié	49	48
<i>Parmi ceux qui qualifient au moins un candidat (N=49):</i>	<i>nb</i>	<i>%</i>
<i>20 % < Tx de qualif < 50 %</i>	7	14
<i>Taux de qualif = 50 %</i>	8	16
<i>60 % < Tx qualif par directeur < 80 %</i>	5	10
<i>Taux de qualif = 100 %</i>	29	60

Sources : données CNU 05. Fichiers candidats et fichiers résultats. Traitements : Sandrine Michel.

III.2.2 Les qualifications et le recrutement de MCF

La mention (parfois encore) obtenue à la thèse ne joue plus le rôle déterminant qu'elle jouait auparavant lors des qualifications des docteurs. En revanche, une règle tacite semblerait s'être imposée : pour être qualifiés, les candidats doivent avoir au moins une publication dans une revue classée Econlit.

La qualification aux fonctions de Maître de Conférences – CNU 05

Tous les candidats				
	2008	2009	2010	2011
Nombre	364	315	338	382
Dossiers non reçus + irrecevables	91	66	83	68
%	25	20,95	24,56	17,80
Nombre de dossier recevables	273			
Qualifiés	132	111	127	170
Taux de qualification	% 48,35	44,58	49,80	54,14
Candidats institutionnalistes				
Nombre	43	60	59	59
Candidats institutionnalistes / total	% 11,81	19,05	17,46	15,45
Dossiers non reçus + irrecevables	13	13	10	13
%	30,23	21,67	16,95	22,03
Dossiers recevables	30	47	49	46
Qualifiés	13	15	18	22
Taux de qualif. institutionnalistes	% 43,33	31,91	36,73	47,83

Sources : données CNU 05. Fichiers candidats et fichiers résultats. Traitements Sandrine Michel.

III.2.3 De MCF à Professeur

L'accession au grade de professeur semble être l'un des points où le resserrement de la discipline s'exprime avec le plus de vigueur. La recherche constitue le principal point d'évaluation, ce qui s'exprime avec une acuité particulière au niveau de l'agrégation externe, la principale voie de promotion. La modification du décret de 1984 dans le sillage de la loi LRU autorise le recrutement des PR sans passage par le CNU pour les étrangers. Nous sommes dans la construction d'un marché du travail hyper concurrentiel ET international.

Quatre voies permettent d'acquérir le grade de professeur :

- L'agrégation externe (art. 46.1). Le concours est accessible à tous les titulaires d'un doctorat. Le nombre de postes ouverts aux autres voies est toujours une proportion des postes de l'agrégation externe.
- L'agrégation interne (art. 46.2). Sur 9 places de professeur mises au concours, deux sont réservées pour ce type de promotion des McF. Cependant, on observera que l'organisation d'un concours d'agrégation interne ne se produit pas à chaque fois qu'est organisé un concours d'agrégation externe.
- La voie longue (art. 46.3). Sur 9 places de professeur mises au concours, une seule est réservée pour ce type de promotion des McF. La qualification ne vaut que pour le(s) poste(s) sur lequel le candidat a été classé. Il vaut donc pour une seule année.
- Autres voies (art. 46.4). Ce type de concours est réservé à des candidats "atypiques". Peu de Pr sont nommés par cette voie dans la mesure où l'article 46.4 cible la

promotion de candidats singuliers, notamment des chercheurs étrangers ou des "professionnels". La qualification est obtenue pour quatre ans. On peut considérer que cette voie de qualification est souvent détournée. Par exemple, une personne d'origine française, jeune, titulaire d'un doctorat obtenu en France, part comme "assistant professor " dans une Université à l'étranger. Elle revient en France au bout de trois ans et se fait embaucher comme professeur associé dans une Université française pendant un an. Elle demande ensuite la qualification par la procédure 46.4, obtient ladite qualification, et le poste de professeur associé est transformé en professeur des Universités.

III.2.4 Recrutement des chargés de recherche et passage de chargé de recherche à directeur de recherche

1) Concours 2009. Pour la section 37, les postes mis au concours sont :

- 2 postes de CR2 dont 1 fléché "gestion des systèmes d'information-gestion"
- 1 poste de CR1
- 4 postes de DR2

Résultats 2009 :

- Postes DR : TSE et PSE
- Poste CR1 : GATE
- Postes CR2 : TSE et GATE + un poste récupéré THEMA

Ils sont tous "mainstream".

2) Concours 2010 : Pour la section 37, les postes mis au concours sont :

- 6 DR2 (dont 1 fléché : "économie comportementale ou formalisée")
- 3 CR2 (dont 1 fléché : "information financière, gestion des risques / gouvernance").
-

Résultats 2010 :

- Postes DR2 : CES ; GREDEG (évolutionniste), un lieu inconnu
- Postes CR2 : École Polytechnique ; CES ; CERSES ; PSE
- Pas de poste CR1 au concours cette année-là

3) Concours 2011 Pour la section 37, les postes mis au concours sont :

- 5 DR2 (dont 1 fléché "théorie économique")
- 1 DR2 économie de l'environnement, (LERNA) à Toulouse,
- 1 CR1
- 1 CR1 fléché sur les thèmes "gestion, marketing, finance"
- 5 CR2 (dont 2 fléchés "économie des réseaux" et "développement durable et espaces géographiques")
- 1 CR2 fléché "énergie, ressources naturelles, climatologie, environnement et développement".

Résultats 2011 :

- Postes DR2 non fléchés : PSE ; Greqam ; CES.
- Poste DR2 fléché : Greqam.
- Poste CR1 non fléché : ?
- Poste CR1 fléché "gestion, marketing, finance" : Pas de candidat classé
- Postes CR2 dont deux fléchés : "économie des réseaux" ; "développement durable et espaces géographiques" : lieu encore incertain, CES (1 en HPE).

En conclusion, les postes ne reflètent pas la diversité de la discipline. Il y a une très forte concentration sur un tout petit nombre de laboratoires. Les candidats en théorie des jeux et en théorie de la décision s'en sortent mieux que les autres.

III.2.5 Les promotions (Hors classe)

L'enseignement et les activités collectives sont marginalement évalués lors les deux étapes clés de la carrière (recrutement et passage au rang A). Les faibles espaces d'évaluation de la mission d'enseignant s'inscrivent dans le cadre du passage en "hors classe".

1) PROMOTION DE GRADE AU SEIN DU CORPS

Promotion des MCF à la hors classe

	2008	2009	2010	2011
Pouvant être promus (1)	nd	nd	367	304
Candidats (2)	147	142	111	85
(2)/(1)			0.30	0.28

Supports promo (3)	14	22	28	30
Promus/Candidats (3)/(2)	0.10	0.15	0.25	0.35

Promus institutionnalistes (4)	5	4	8	7
Promus institutionnalistes / Promus (4)/(3)	0.36	0.18	0.29	0.23

Sources : données CNU 05. Fichiers candidats et fichiers résultats. Traitements : Sandrine Michel.

Nota : en 2010 changement pour les promotions : début de la procédure Electra

	MCF (supports promo hors classe)	PR (supports promo)
2008	147 (14)	PR1 : 100 (9) PRCE 1 : 65 (9) PRCE 2 : 17 (4)
2009	142 (22)	PR1 : 92 (12) PRCE 1 : 91 (8) PRCE 2 : 19 (4)
2010	111 (28)	PR1 : 85 (11) PRCE 1 : 82 (13) PRCE 2 : 13 (3)
2011	85 (30)	PR1 : 72 (14) PRCE 1 : 61 (13) PRCE 2 : 17 (5)

QUELQUES ÉLÉMENTS DE BILAN À PROPOS DES PASSAGES DES MCF À LA HORS CLASSE

Une auto sélection dommageable :

- Moins du 1/3 des EC pouvant être promus déposent un dossier de candidature.
- Le passage à la procédure Electra a sans doute découragé nombre de collègues par sa lourdeur. Mais en aucun cas, Electra n'a créé le problème de l'auto-sélection.
- Les résultats finaux plaident contre cette auto sélection.

Évaluation CNU :

- Le CNU évalue les 3 volets suivants : recherche + pédagogie + responsabilités collectives (critères sur le site de la CP-CNU section 05). Chaque volet est évalué A, B, C.
- Aucune "note" globale n'est attribuée.
- Un C disqualifie en général, mais certains C pèsent plus que d'autres. Ainsi, un C recherche est rédhibitoire, tandis qu'un C en responsabilités collectives ne l'est pas.
- Il n'est pas facile de dire quel est le profil du candidat promu. Il y en a sans doute plusieurs. Mais les profils équilibrés ont été correctement traités au cours de cette mandature (...).

Le nombre de supports

- Bien entendu, les sessions sont plus ou moins tendues selon le nombre de supports mis à la disposition de la section.
- Depuis 2009, les sessions de promotion des MCF se sont de mieux en mieux passées.

Le retour au local

- Depuis 2010, le CNU rend un avis qui retourne vers les Universités. La question est de savoir ce qu'elles en font. Pour l'instant, la réponse est assez claire : rien...

Les promus institutionnalistes

- Pas de discrimination évidente à leur égard.
- Sont très engagés dans le fonctionnement de l'Université.

Le CNU partage les possibilités de promotion avec les Universités. Depuis 2010, la procédure Electra rend obligatoire les avis des CA restreints sur la pédagogie et les responsabilités collectives en direction du CNU, mais le CNU doit également faire redescendre ses avis vers les Universités. Un bilan a été dressé sur les promotions par la voie locale pour l'année 2010. Il met en évidence l'existence de deux modèles clairement différenciés de promotion.

Promotion des PR - Campagne 2010 - Suivi des avis CNU

Avis CNU	Nombre (1)	Promotions nationales (2)	Promotions locales (3)	Observations (3)/(1)
Avis 1	27	27	-	
Avis 2	35	Sans objet	9	25,7 %
Avis 3	108	Sans objet	22	20,3 %
Avis 4	4	Sans objet	0	0 %
Dossiers non reçus	6			
Total	180	27	31	

Promotion des MCF - Campagne 2010 - Suivi des avis CNU

Avis CNU	Nombre (1)	Promotions nationales (2)	Promotions locales (3)	Observations
Avis 1	28	28	-	
Avis 2	21	Sans objet	8	38%
Avis 3	42	Sans objet	20	47,6%
Avis 4	19	Sans objet	4	21%
Total	110	28	32	

(1) Nombre d'avis 1, 2, 3, 4 émis par la section du CNU

(2) Nombre de promus CNU avec avis 1

(3) Nombre de promus locaux par type d'avis CNU : par ex, n promus locaux ayant obtenu un avis 2 ; n promus locaux ayant obtenu un avis 3, etc.

Avis 1 = promotion nationale

Avis 2 = aurait été promu si contingent national l'avait permis (la section 05 a décidé d'attribuer cet avis au maximum au nombre de candidats promus)

Avis 3 = un volet du dossier est insuffisant

Avis 4 = dossier insuffisant

III.3. Les propositions de l'AFEP

Discourir sur la pertinence de l'évaluation des EC ne constitue pas l'objectif de ce groupe de travail, elle est un fait et si elle est maintenue, elle doit mettre en valeur toute l'activité scientifique et pédagogique.

L'excellence (si ce mot a un sens dans le contexte de recherche) ne peut se juger à l'aune des critères établis par la commission 37. La liste retenue privilégie un mode de pensée, et pénalise le risque de s'aventurer sur des voies inconnues et innovantes.

Les mécanismes et dispositifs de l'excellence chassent la majorité de la profession.

- Enjeu de l'existence même de la profession.

- Enjeu d'un rétrécissement de la discipline, voire d'une "clanification"
 - Reconnaissance de la diversité des missions, des carrières, i.e. les carrières basées sur l'investissement pédagogique ou collectif.
 - Reconnaissance de cette diversité y compris lors des changements de grade.
 - La fécondité de la recherche, quel que soit le domaine, repose sur une lecture critique des faits et des idées, sur la diversité des modes et des écoles de pensées. En d'autres termes l'évaluation, doit s'inspirer d'abord du pluralisme.

Les conséquences néfastes de l'actuelle façon d'évaluer l'activité des enseignants-chercheurs sont nombreuses et considérables, notamment pour ce qui est de l'innovation scientifique. L'AFEP propose de reconsidérer profondément le mode d'évaluation des enseignants-chercheurs en reconnaissant la diversité de leurs missions, de leurs activités et celle de leurs pratiques. Plutôt que de les considérer comme des individus isolés et calculateurs, intéressés de façon principale par l'optimisation de leur carrière et leur accès aux primes, l'AFEP entend promouvoir la liberté du chercheur et son autonomie dans le cadre de collectifs orientés par le souci d'une progression de chacun et de tous, au service de la société.

Les propositions de l'AFEP pour ce qui est de l'évaluation des enseignants-chercheurs se déclinent autour de six axes.

Ces axes sont par la suite déclinés dans la quatrième partie de ce document (voir préconisations).

Les six axes sont les suivants :

1. Prise en compte et valorisation de l'ensemble des activités : l'enseignement et la pédagogie au même titre que recherche.
2. Évaluation de l'activité de recherche et non réduction à tel indicateur de performance.
3. Prise en compte des temporalités de la carrière et du caractère irrégulier des productions et résultats scientifiques.
4. Encouragement des pratiques coopératives par la prise en compte des responsabilités collectives de toutes sortes (coordination de contrats de recherche collectifs, coordination de publications, participation à des comités éditoriaux, participation à la direction de masters, de laboratoires, etc.) ainsi que par l'encouragement aux pratiques collectives de recherche et de publication (recherche en réseau, co-publications, etc.).
5. Des évaluations moins fréquentes (moins nombreuses ?) mais plus approfondies, davantage tournées vers l'amélioration des pratiques.
6. Suppression des concours d'agrégation pour le recrutement des Professeurs des Universités et reconnaissance du diplôme d'habilitation à diriger des recherches comme ouvrant vers le grade de Professeur.

IV. Synthèse et préconisations

Synthèse de : Gabriel Colletis et Thomas Lamarche

Contributions de : Pierre Alary, Florence Audier, Jean-Pierre Chanteau, Hervé Charmettant, Gabriel Colletis, Antonella Corsani, Ai-Thu Dang, Florence Jany-Catrice, Thomas Lamarche, Catherine Laurent, Philippe Méhaut, Sandrine Michel, Martino Nieddu, Christian du Tertre, Serge Walery.

Cette partie entend d'abord proposer une synthèse centrée sur l'explicitation du mode d'évaluation que l'AFEP entend promouvoir. Cette synthèse reprend et rassemble des éléments qui ont été exposés dans les parties précédentes.

Par la suite, des préconisations sont faites aux deux niveaux d'évaluation appréhendés dans ce document : celui des laboratoires et celui des enseignants-chercheurs.

IV.1. Pour une évaluation professionnelle et non une évaluation sanction

IV.1.1 *S'assurer a minima du respect des règles et des procédures*

Une condition tout à fait nécessaire mais largement insuffisante dans la perspective d'une évolution des pratiques d'évaluation consisterait à procéder à des évaluations en utilisant des critères précis, transparents et explicites, et selon des procédures elles-mêmes claires et explicites.

Une réflexion sur l'évaluation viserait à *instaurer* des procédures rendant les règles d'évaluation transparentes et à obliger au respect de ces procédures. Un travail qui viserait uniquement à préciser une liste de critères d'évaluation serait un premier pas, mais une absence de précision sur les procédures le rendrait vain.

Pour chaque type d'évaluation, la transparence serait exigée concernant : a) les objectifs des activités évaluées (politiques explicites d'enseignement et recherche); b) l'objet de l'évaluation (le périmètre des tâches, des domaines de résultats...) et c) les procédures de contrôle (la forme des rapports, le statut des observateurs, les modalités de recours, ...).

Ce qui précède prend sens si ces principes généraux édictés pour chaque item :

- sont traduits en *critères explicites* d'évaluation, avec éventuellement des pondérations,
- sous forme d'une *liste exhaustive* des critères qu'il est légitime de prendre en compte pour l'évaluation.

Ce qui précède prend sens aussi si les évaluateurs rendent compte obligatoirement de *tous* les critères d'évaluation dans leur délibération et leur rapport (et encore plus si ces délibérations et ces rapports s'effectuent sous l'œil d'un observateur indépendant).

Une seconde condition impérative consisterait à ce que les critères d'évaluation soient discutés par l'ensemble de la communauté, et constituent une convention partagée, ce qui n'est absolument pas le cas actuellement, comme on l'a montré *supra*.

Ces principes, très simples, ne permettent pas de régler – tant s'en faut – la totalité des problèmes. Compte tenu de la situation actuelle, ces principes ont cependant un certain potentiel émancipateur. De surcroît, ils correspondent aux standards européens qui peuvent être difficilement récusés par nos institutions.

IV.1.2 Accompagner et non vérifier la conformité à une norme

Mais comme cela a été dit, une simple évolution dans le sens du respect des règles et de la transparence des modalités d'évaluation est très loin de suffire.

L'évaluation ne doit pas être dogmatique, conduite pour vérifier la conformité à une norme privilégiant telle approche plutôt que telle autre, telle méthode, tel ou tel objet, tel ou tel support.

En cohérence avec les missions qu'elle s'est fixées, l'AFEP défend les processus d'évaluation contradictoires et pluriels, et souhaite promouvoir une évaluation respectant le pluralisme des objets, des méthodes et des approches conceptuelles. La recherche en économie ne peut se réduire à une conception où la pratique de la recherche est réduite à un ensemble de techniques (qui tend même à se réduire à des techniques quantitatives) qu'il faudrait maîtriser, alors même que l'AFEP¹⁵ défend l'idée que l'économie doit toujours être définie historiquement, que la boîte à outils conceptuels est riche de ce que tout ce que lui a légué l'histoire de sa discipline. L'AFEP considère aussi que les modes d'administration de la preuve sont multiples, empruntant au bagage commun des autres sciences sociales.

L'AFEP s'oppose à une évaluation qui dérive vers une pratique de contrôle, et qui ne conçoit la recherche que comme l'application de ces techniques, assimilant la non-soumission à ces techniques à l'absence de travail, ou de résultats.

Pour l'AFEP, l'évaluation a pour but de marquer d'éventuelles étapes dans le développement de programmes et d'engagements, et de fournir aux évalués un regard extérieur, et accompagnant. Le changement profond du régime d'évaluation que l'AFEP entend promouvoir s'inscrit dans cette perspective d'accompagnement des laboratoires comme des enseignants-chercheurs. Cela inscrit l'exercice d'évaluation dans le temps long.

Dans cette perspective, il convient de veiller à ce qu'il n'y ait ni conflits d'intérêts, ni défiance réciproque entre évaluateurs et évalués, et à proscrire *in fine* toute pratique de notation, quota et classement. Les notations tendent en effet à doter l'évaluation d'une illusoire objectivité, quasi-mécanisable, et produite rapidement. L'AFEP pense au contraire que les processus d'évaluation contradictoires et pluriels qu'elle revendique nécessitent de prendre du temps.

Le changement profond du régime d'évaluation que l'AFEP entend promouvoir s'inscrit dans une perspective d'accompagnement des laboratoires comme des enseignants-chercheurs.

Cet accompagnement inscrit l'exercice d'évaluation dans le temps long.

L'évaluation professionnelle que l'AFEP entend promouvoir considère en premier lieu non les "résultats," mais les pratiques professionnelles et les capacités collectives et individuelles à sortir des "modes", à faire avancer les pratiques pédagogiques et de production de connaissances à caractère novateur : le fait d'avoir "décroché" une ANR ou encore un LabEX, ou d'avoir réussi à publier dans telle ou telle revue ne peut se substituer à ces critères, ou en tenir lieu.

L'évaluation doit tenir compte des ressources dont disposent les équipes pour mener à bien leurs travaux. Elle doit aussi concerner les travaux conduits dans le cadre de programmes, de

¹⁵ Voir le groupe de travail sur "Pour une nouvelle section".

conventions et d'autres contrats financés par les deniers publics (conventions de recherche, ANR, Labex, etc ...).

Cette position générale conduit à questionner l'existence de l'AERES, car l'AFEP n'entend pas que perdurent des pratiques que l'AERES véhicule. Véritable "agence de notation", l'AERES s'oppose en l'état aux pratiques évaluatives que l'AFEP souhaite promouvoir. Pour les mêmes raisons, l'AFEP refuse également la constitution, *de jure* comme *de facto*, d'un corps de professionnels de l'évaluation. L'évaluation devrait être considérée comme faisant normalement partie des pratiques professionnelles de chacun.

IV.1.3 Pour une évaluation fondée sur le contenu réel du travail d'enseignement et de recherche et non sur un régime d'évaluation quantitative.

L'AFEP dénonce le régime actuel d'évaluation à caractère exclusivement quantitatif. Dans le domaine de la recherche, seul compte désormais le nombre de publications de rang 1* et 1. Les laboratoires, considérés non pour leur projet scientifique mais comme une somme de chercheurs produisant, voient leurs moyens calculés sur cette seule base.

L'AFEP refuse le régime d'évaluation quantitatif pour les enseignants-chercheurs comme pour les laboratoires (UMR et équipes d'accueil). La quantification de la qualité d'un EC ou d'un laboratoire est une fiction qui vise *in fine* à réduire la variété des activités de recherche à une seule de ses composantes : la publication dans des revues de rangs 1* et 1. L'AFEP refuse le principe de mise en concurrence des chercheurs comme des laboratoires et s'oppose à la constitution d'un "marché" de la recherche, tant pour les enseignants-chercheurs que pour les laboratoires.

Ce refus de la concurrence est conjoint d'une très grande réserve à l'égard de la notion d'excellence. L'AFEP considère en effet que le périmètre de la recherche en économie doit être le plus étendu possible pour faire face à des défis socio-économiques majeurs. Or, la logique d'excellence qui est mise en place se traduit d'abord par un écrémage et par une diminution préjudiciable de la diversité des recherches menées. En effet, la hiérarchisation des activités et des revues incite les chercheurs et les laboratoires vers les activités les plus "rentables" au regard des critères imposés.

L'AFEP se prononce pour une évaluation portant sur le contenu réel du travail d'enseignement et de recherche, et encourage le développement des pratiques coopératives (entre enseignants-chercheurs, chercheurs, et entre équipes de recherche). L'enseignement comme la recherche doivent être orientés vers les savoirs utiles, qu'il s'agisse de leur diffusion comme de leur production. Les savoirs utiles sont les savoirs orientés par les débats et les besoins de la société, et ceux qui participent au progrès général des connaissances, et à l'émancipation des individus.

L'AFEP entend renforcer, voire, dans certains cas, rétablir la collégialité. La collégialité et le pluralisme dans la vie des sociétés savantes sont vitaux pour la démocratie comme pour la production et la transmission des savoirs. La délibération contradictoire, la responsabilité collégiale, l'évaluation des évaluateurs sont des facteurs essentiels d'une évaluation scientifique rigoureuse, libre de considérations idéologiques et financières, non soumise aux autorités de toutes sortes : administratives, politiques, économiques.

IV.2. Préconisations concernant l'évaluation des laboratoires de recherche

En cohérence avec sa conception de l'évaluation, et à l'instar du CoNRS, l'AFEP entend que soit mis fin à la "notation" qui ne peut que déboucher sur une distribution de "bons" et "mauvais" points, donc sur un classement et, au final, des sanctions (fermeture de laboratoires, désassociations, baisse de la dotation).

L'AFEP dénonce le fait que les jugements ainsi prononcés et rendus publics soient sans appel. Elle rappelle les conséquences de ces jugements : ils cristallisent des situations pour 4 ou 5 ans, ce qui – outre le caractère injuste et les effets de renommé qui y sont attachés – est contraire à une démarche d'évaluation positive.

L'AFEP estime que toute évaluation d'un laboratoire de recherche doit partir du rappel de ses missions :

élaboration de questions de recherche ayant une finalité d'avancée des connaissances et/ou de production de savoirs utiles ;

- production de dispositifs de réponse à ces questions mobilisant les différentes dimensions de la recherche : épistémologique, théorique, conceptuelle, méthodologique, empirique ;
- formation des chercheurs par et pour la recherche (dont accueil des doctorants et post-doctorants) ;
- fertilisation de l'enseignement par les méthodes et résultats de la recherche (dont articulation avec les masters) ;
- diffusion des connaissances et du savoir vers des publics différenciés.

L'AFEP propose que la conduite des évaluations :

- reflète la pluralité des missions des laboratoires ;
- ait pour finalité d'aider à l'affirmation de leur projet scientifique ;
- garantisse, par le caractère démocratique de leur élaboration et de leur utilisation, le pluralisme des approches, des objets de recherche, et des méthodes ;
- exprime de façon privilégiée la dimension collective de l'organisation et du fonctionnement de la recherche ;
- favorise l'ouverture disciplinaire.

L'évaluation pluraliste des laboratoires proposée par l'AFEP recouvre les items suivants :

IV.2.1 La prise en compte la dimension longue du temps de la recherche

La recherche ne doit pas être au service exclusif de besoins préalablement identifiés et immédiats. La dimension longue du temps est essentielle dans la dynamique de recherche. Dans cette perspective, les moyens alloués à la recherche doivent être orientés vers l'activité scientifique des laboratoires, et non guidés par des objectifs intermédiaires comme l'activité de publication.

Lorsque les politiques scientifiques sont orientées vers l'interdisciplinarité, et ce sont des recommandations y compris en SHS, alors il faut aussi prendre en considération les tentatives

faites en ce sens, même si elles ne débouchent pas immédiatement. La construction d'objets dans l'interdisciplinarité nécessite du temps.

IV.2.2 L'évaluation par les pairs

Le principe de lecture et d'appréciation des travaux induit, inévitablement, celui d'évaluation par les pairs. Or, le passé de l'Université française a abondamment montré que l'évaluation par les pairs pouvait être, elle aussi, profondément conformiste et génératrice de biais. Il importe donc de ne pas se contenter d'un simple retour au principe d'évaluation par les pairs, mais d'organiser et de promouvoir systématiquement la diversité et le pluralisme, tant des instances qu'au sein des instances.

Le jugement collégial d'une communauté de pairs devrait donc être privilégié. Ce jugement devrait reposer sur trois principes. Le principe d'autonomie : les règles doivent être définies par ceux qui les mettent en œuvre ; le principe d'indépendance : les règles doivent procéder des professionnels et non des autorités de tutelle ; le principe de pluralisme : pluralisme des approches, des méthodes, des objets.

IV.2.3 Le principe essentiel de lecture des travaux

Enrayer la dérive quantitativiste du régime d'évaluation suppose, en premier lieu, que la substance du pouvoir d'évaluation ne soit plus accaparée par les revues et les listes, mais reprise par les instances qui en sont officiellement dotées. Or, une telle ré-appropriation ne peut être envisagée que si les instances prévues à cet effet en reviennent à une évaluation sur travaux (et non sur indicateurs). La dérive quantitativiste ne saurait être combattue sans que soit systématiquement réaffirmé le principe selon lequel toute évaluation sans lecture et appréciation des travaux devrait être proscrite.

De ce principe élémentaire découlent, concrètement, de multiples effets ou conséquences. La lecture et l'appréciation des travaux (ou d'une sélection de travaux) permettraient, tout d'abord, de rendre leur importance aux formats autres que l'article, laissant ainsi plus de place à des démarches moins "standard".

Pour soutenir aujourd'hui les progrès de la connaissance scientifique, et spécifiquement pour œuvrer au renouvellement de l'économie, il convient enfin de mener le combat du vocabulaire, beaucoup moins anodin qu'il n'y paraît. Face aux impasses de l'évaluation quantitative de la productivité académique, il importe de promouvoir, inlassablement, une logique d'appréciation de la contribution scientifique des chercheurs et des laboratoires. Cette logique d'appréciation recouvre les questions de langage et de sémantique, qui, le plus souvent, reflètent et résument des questions tout à la fois épistémologiques, théoriques, conceptuelles et méthodologiques qui ne peuvent s'exprimer seulement dans un univers de quantification.

Sont aujourd'hui déconsidérées les revues de rang 3 et 4, quel qu'en soit le contenu, surtout si l'article est écrit en français... ce qui est le cas des revues françaises dont aucune n'accède au rang 1.

La promotion des travaux en langue française n'est en rien un combat d'arrière-garde, c'est un combat sur le sens. L'expression n'est en rien disjointe du contenu de ce qui est exprimé. De plus, dès lors que l'accent est mis sur le lien enseignement/recherche, la production de textes en français s'impose si on veut être lu par un public étudiant largement français ou francophone.

Concernant la "production", il s'agit d'accepter une certaine variété de publications, lesquelles ont des "cibles" différentes. Les revues académiques, mais aussi les ouvrages ou chapitres d'ouvrages, les comptes rendus de colloques, et même les rapports de recherche. Lorsqu'existaient les GIP (Groupements d'Intérêt Public), par exemple, ou les ATP (Actions thématiques programmées) du Cnrs, ce qui était attendu et évalué, c'étaient les rapports de recherche, indûment appelés "littérature grise". Il y a bien une différence entre les études et les recherches, et ce n'est qu'à la seule lecture des documents produits qu'on peut percevoir cette différence.

IV.2.4 L'usage à modérer fortement des indicateurs bibliométriques

Parallèlement, récuser toute évaluation sans lecture des travaux imposerait, matériellement, une réduction significative du nombre et de la fréquence des évaluations, laquelle favoriserait la qualité effective de l'évaluation, et limiterait, au moins à court terme, les incitations à la sur-publication. L'application de ce principe permettrait, selon les préconisations de l'Académie des Sciences (voir annexe 12, Académie des Sciences, 2011) d'interdire le recours aux indicateurs bibliométriques pour l'évaluation des jeunes chercheurs, et d'en circonscrire étroitement l'usage pour les autres.

IV.2.5 Le soutien de thématiques de recherche innovantes

L'innovation, les recherches orphelines et/ou hétérodoxes devraient être favorisées. Parallèlement, la logique d'examen devrait être systématiquement privilégiée, la logique de concours n'intervenant qu'en second temps, et seulement dans les cas où elle se révèle indispensable.

IV.2.6 La prise en compte des apports méthodologiques

L'activité scientifique ne peut pas se résumer aux textes écrits et aux publications : il faut aussi prendre en considération les apports méthodologiques ainsi que la construction de nouveaux outils de connaissance, par exemple la construction de bases de données. Si celles-ci sont exclusivement déléguées aux grandes institutions comme l'INSEE ou Eurostat, alors c'est que l'on considère que la recherche est découplée de ses outils. Or, on le sait, les outils sont loin d'être axiologiquement neutres (ce débat rejoint la question du vocabulaire et de la sémantique).

IV.2.7 Un soutien privilégiant les laboratoires en difficulté ou en démarrage

Un soutien accentué des laboratoires en phase de démarrage (jeunes équipes) ou à des laboratoires en difficulté peut impliquer une répartition des moyens accordant à ces laboratoires une dotation plus que proportionnelle à leur taille ou à leur rayonnement scientifique.

IV.3. Préconisations concernant l'évaluation des enseignants-chercheurs

Le schéma ci-après se propose de partir de l'analyse de l'évaluation telle qu'elle est en train de se déployer dans l'Enseignement supérieur et la recherche et de lui opposer une évaluation alternative.

Rappelons que la discussion de l'évaluation est une discussion qui se déploie sous un double aspect :

- l'AFEP est favorable à l'évaluation. En effet,
 - ♦ En tant qu'agents publics, les enseignants-chercheurs comme les chercheurs doivent des comptes sur l'exercice de leurs missions professionnelles.
 - ♦ L'évaluation a été revendiquée par les personnels dont une partie croissante de la charge de travail est méticuleusement rendue invisible.
- l'évaluation en soi est utile. Ce sont finalités et les modalités d'exercice actuelles que l'AFEP conteste.

Le schéma ci-après est centré sur les deux modes d'évaluation des enseignants-chercheurs, le premier (actuel) que nous contestons, le second que l'AFEP propose.

Comme on le sait, les indicateurs de performance se déclinent, en effet, aux niveaux individuel et collectif ; les palmarès et les classements orientent les politiques publiques comme les stratégies d'établissement ou les conduites individuelles ; les conventionnements sur projet, tributaires des évaluations, retentissent directement sur les pratiques de recrutement et de rémunération (Annexe 5, M.-C. Bureau).

De l'« évaluation sanction de l'EC » actuelle...

- centrée sur les résultats/performances de la seule activité de recherche
- guidée par un principe de mise en concurrence et d'élitisme

Principes d'action : norme excluante

- Contrôle de l'accomplissement de la mission R
 - Autorenforcement : mêmes critères \forall instance
- ↔ **Dimension transversale = mise en concurrence**

Objets :

- Individu (Modulation : mission R contre mission E)
- Collectifs de recherche (Attribution de moyens basée sur excellence)



...à une « évaluation promotion des activités de l'EC » demain

- s'intéressant aux moyens mis en œuvre pour l'accomplissement des missions
- intégrant les dimensions de coopération valorisant le travail collectif

Principes d'action : norme non excluante + conseil

- Contrôle de l'accomplissement des missions (norme/moyens)
 - Evaluation des contenus (conseil)
- ↔ **Dimension transversale = coopération**

Objets :

- Individu (sans modulation)
- + Collectifs (Attribution de moyens inclusive)
- Individus et collectifs les plus coopératifs sont valorisés : publications (toutes) + partages d'info et de procédures

IV.3.1 L'évaluation aujourd'hui : l'évaluation sanction

L'évaluation mise en place actuellement a un double objectif :

- c'est un outil de gestion visant à élever la productivité des personnels dans un contexte de pénurie durable de moyens ;
- c'est un outil de contrôle de la profession dans l'exercice de ses missions mais également des contenus des recherches.

Ainsi, l'évaluation se veut l'énoncé normatif d'une vérification et la légitimation d'une sanction.

Chacun ici doit périodiquement rendre des comptes permettant de trier entre ceux qui se conforment à une norme établie pour la recherche, conçue comme excluante, et qui va influencer l'exercice des autres missions, et les autres.

La fabrication de la norme privilégie une mission, la recherche, et se concentre sur un énoncé réducteur de la production en recherche. Décidée à peu, elle exprime l'exercice d'un pouvoir d'auto-sélection. Elle se déploie selon un petit nombre de critères opérationnels, adoptés de manière implicite par toutes les instances d'évaluation, contribuant en cela à une diffusion généralisée à la profession et à l'exercice de toutes les missions des personnels. C'est ce que nous désignons sous le terme d'auto-renforcement par l'action de méta-règles.

Son avantage est qu'elle est vérifiable par tous. Cela lui confère une apparente légitimité. Sensible au rapport de forces, elle est susceptible d'enregistrer des modifications plus ou moins importantes, tout en maintenant pleinement actif son objectif général.

Cette approche normative se déploie sur la base d'une mise en concurrence des individus et des collectifs de recherche. Dans une application triviale de la théorie des marchés, ceux qui sont capables de payer le prix, dans le sens où ils remplissent les critères (individus ou collectifs de recherche) vont concentrer une grande partie des moyens et verront l'accomplissement de leur mission de recherche facilitée. Ceux qui ne peuvent payer le prix, sont exclus. Pour eux, la sanction se déploie avec l'impossibilité d'exercer l'une de leur mission, ou l'obligation de disparaître en tant que collectif de recherche. L'avantage est ici purement individuel. Il instrumente un élitisme soigneusement défini par ceux-là même qu'il vise à protéger.

Ce mode d'évaluation développe au final une véritable ingénierie du soupçon prospérant dans un univers professionnel se nourrissant de la division des personnels et constitue une négation de l'une des missions fondamentale de l'Université et donc de ses personnels : la formation.

IV.3.2 L'évaluation professionnelle

Nous proposons de substituer à cette évaluation sanction une évaluation professionnelle visant non à sélectionner, classer, hiérarchiser les enseignants-chercheurs ou les chercheurs, mais à les aider dans la promotion de leurs objectifs librement élaborés.

Cette évaluation alternative poursuit plusieurs objectifs :

- le contrôle de l'accomplissement des missions, avec une vérification qui lie explicitement les résultats obtenus aux moyens réellement mis en œuvre. Ceci revient à dire que l'objet de l'évaluation des enseignants-chercheurs n'est pas d'évaluer ce qu'ils sont mais ce qu'ils font ;
- un examen approfondi des contenus des recherches et des pratiques pédagogique, visant le soutien mais aussi le conseil.

Cette proposition alternative repose sur notre expérience professionnelle : nul ne peut prétendre chercher, trouver et publier seul. Aujourd'hui, la recherche est une activité collective ; elle repose sur des synergies. Nous nous servons de cette caractéristique fondamentale de la production scientifique pour avancer que la coopération procure des bénéfices nets supérieurs à toute démarche individuelle si la condition de confiance et la production de normes propre à la communauté respecte les projets de chacun (Ostrom).

Les individus et les collectifs de recherche doivent donc remplir les missions que leur confère leur statut ou le contrat de leur création. L'établissement d'une norme vérifiable est légitime, à condition qu'elle procède d'un processus démocratique. Ainsi, dans l'esprit de notre proposition, tous les types de publications doivent être pris en considération dans les évaluations individuelles et collectives.

Les individus et les collectifs ne sont pas des "machines" à gagner. Tout programme de recherche connaît des hauts et des bas. Nous considérons comme particulièrement efficace qu'une phase basse, de difficulté, ne soit pas automatiquement sanctionnée par un abandon. Mais qu'elle donne au contraire lieu, de manière volontaire ou pas, à un processus d'évaluation par des pairs aboutissant à des recommandations visant à ce que les acquis, et il y en a toujours, soient redistribués pour des réorientations productives.

En adéquation avec nos pratiques professionnelles et en connaissance de cause des dynamiques des programmes de recherche ou de formation, nous pensons que les individus et

les collectifs les plus coopératifs doivent être les plus valorisés. Cette proposition vise une efficacité supérieure pour l'Enseignement supérieur et la recherche comme pour la société dans son ensemble.

En résumé, la base sur laquelle l'évaluation des enseignants-chercheurs devrait être conduite est la suivante :

1. *Prise en compte et valorisation de l'ensemble des activités : enseignement au même titre que recherche*
 - a. Refus d'une hiérarchisation systématique des activités.
 - b. Refus de tout alourdissement des services d'enseignement. L'enseignement n'est pas une peine ou une sanction. Il doit être valorisé.
 - c. Suppression des primes dites d'excellence (donc de la PES) et de toutes primes non liées à des activités spécifiques. Refus du principe que la performance puisse être considérée comme individuelle.
 - d. Maintien de primes liées à des activités spécifiques (encadrement d'étudiants, responsabilités de formations, tâches de direction de laboratoires, etc.), sans que cela serve de motif pour refuser une revalorisation générale des traitements de base des enseignants-chercheurs.

2. *Évaluation de l'activité de recherche et non de tel indicateur de performance*
 - a. Lecture effective des productions scientifiques et non simple comptage¹⁶
 - b. Prise en compte la plus large possible de tous les types de publication : articles, chapitres d'ouvrages et ouvrages, rapports publics, etc., en fonction des projets et programmes scientifiques.
 - c. Prise en compte des autres formes de "production" comme les enquêtes, les bases de données etc., dès lors qu'elles sont scientifiquement validées.
 - d. Prise en compte des activités mixant formation et recherche : encadrement, production de manuels et de documents pédagogiques...
 - e. Prise en compte des pratiques de diffusion de la recherche. De l'action dans le débat public.
 - f. La défense de la pluralité des démarches et des paradigmes est l'enjeu central de l'économie politique :
 - i. Pluralisme en économie / pluralisme des représentations du monde
 - ii. Pluralisme des approches conceptuelles
 - iii. Pluralisme des points de vue
 - iv. Pluralisme des objets
 - v. Pluralisme des méthodes
 - vi. La question cruciale de la construction des "données"
 - vii. Ouverture aux autres disciplines¹⁷.

¹⁶ Pour l'Académie des Sciences, seule l'évaluation qualitative peut apprécier l'originalité et l'esprit d'innovation d'un chercheur. Cette réserve majeure implique que la bibliométrie ne doit être utilisée que par les pairs, et surtout pas par des personnes n'en ayant pas l'expertise, ou n'ayant pas pris le temps nécessaire pour examiner les dossiers scientifiques dans leur ensemble (Annexe 12, Académie des Sciences, 2012).

3. *Prise en compte des temporalités de la carrière et du caractère irrégulier des productions et résultats scientifiques*
 - a. Prise en compte de la parité pour limiter les effets déstructurant des carrières des femmes.
 - b. Possibilité donnée aux jeunes enseignants-chercheurs et aux enseignants-chercheurs expérimentés qui le souhaitent et qui en font la demande d'obtenir un allègement transitoire de leur service d'enseignement afin de leur permettre de franchir une étape en matière de recherche.

4. *Encouragement des pratiques coopératives* par la prise en compte des responsabilités collectives de toutes sortes (coordination de contrats de recherche collectifs, coordination de publications, participation à des comités éditoriaux, participation à la direction de masters, de laboratoires, etc.) ainsi que par l'encouragement aux pratiques collectives de recherche et de publication (recherche en réseau, co-publications, etc.).
 - a. Refus de la logique de l'excellence individuelle, comme de l'excellence des laboratoires qui renforcent l'écroulement et la concurrence

5. *Des évaluations moins fréquentes mais approfondies, davantage tournées vers l'amélioration des pratiques.* Dans le cadre d'une évaluation visant à l'amélioration des pratiques, et afin de favoriser la diversité des perspectives dans le processus d'évaluation, fournir aux instances d'évaluation les moyens de consulter (oral/écrit) : les spécialistes du domaine, les collègues du laboratoire de recherche, ceux du département d'enseignement, (les anciens étudiants ?), etc. Une synthèse de ces consultations serait naturellement remise (et expliquée) à l'EC concerné.

6. *Suppression des concours d'agrégation* pour le recrutement des Professeurs des Universités et *reconnaissance du diplôme d'habilitation à diriger des recherches* comme ouverture vers le grade de Professeur.

¹⁷ La section 37 a ajouté à sa liste de revues "classées" une liste de revues non classées (nommée MAD), qui représentent des revues dans lesquelles les économistes publient ou ont publié. Voir annexe 15

V. Annexes

<i>Annexe 1</i>	59
L' AERES et le mythe de la "masse critique"	59
<i>Annexe 2</i>	61
Interprétation des notations de l' échelle de notation multicritères	61
<i>Annexe 3</i>	63
L' évaluation contre la science ; productivité académique vs contribution scientifique : le cas de l' économie	63
<i>Annexe 4</i>	72
La loi relative aux libertés et responsabilités des Universités : essai de mise en perspective	72
<i>Annexe 5</i>	73
Du travail à l' action publique : quand les dispositifs d' évaluation prennent le pouvoir.....	73
<i>Annexe 6</i>	75
Réflexions croisées sur la mesure et l' incertitude	75
<i>Contribution de François Vatin : L' incertitude du monde social et la raison métrologique : une approche pragmatiste</i>	75
<i>Contribution d' Alain Caillé : Petit plaidoyer pour une incertitude modérée et une mesure mesurée</i>	76
<i>Contribution d' Olivier Favereau : L' incertitude sans mesure</i>	77
<i>Annexe 7</i>	79
Revue qui compte, revue qu' on compte : produire des classements en économie et gestion. 79	
<i>Annexe 8</i>	82
External review report of the French Evaluation Agency for Research and Higher Education (AERES) with regard to the Standards and Guidelines for Quality Assurance in the European Higher Education Area, May 2010.	82
<i>Annexe 9</i>	86
Le danger des mauvais usages des indicateurs	86
<i>Annexe 10</i>	87
Réflexion sur l' évaluation et les listes de revues. L' exemple de l' économie et de la gestion.....	87
<i>Annexe 11</i>	90
Une étude comparative internationale des bases de données des revues scientifiques en sciences humaines et sociales (SHS)	90
<i>Annexe 12</i>	91
Rapport de l' Académie des Sciences Du bon usage de la bibliométrie pour l' évaluation individuelle des chercheurs.....	91
<i>Annexe 13</i>	93
Quelques enseignements statistiques de l' évaluation des laboratoires par l' AERES	93
<i>Annexe 14</i>	99
Pour un débat sur les représentations en économie	99
<i>Annexe 15</i>	103
Une nouvelle catégorie de revues "multidisciplinaires ou d' autres disciplines" (MAD)	103

Annexe 1

L'AERES et le mythe de la "masse critique"

Grossetti M., Milard B., Eckert D., Bouba-Olga O.

"L'agence d'évaluation et la géographie des sciences", (2011)

<http://www.scribd.com/doc/52118695/Deconcentration-de-la-recherche>,

accès le 4/5/2011 (cité in <http://obouba.over-blog.com/article-labex-masse-critique-et-autres-futilites-70800571.html>).

-Extraits-

Selon Jean-François Dhainaut, directeur de l'AERES (cité par *Les Echos.fr*, 11.01.11), "pour un certain nombre de régions hors des grandes métropoles, des collaborations serrées seront indispensables pour atteindre la masse critique qui leur permettra de travailler à armes égales" exprime une idée très répandue dans les instances de pilotage de la recherche et que l'on peut baptiser "la théorie de la masse critique". L'expression "masse critique" fait référence de façon métaphorique à la masse de matériau radioactif à partir de laquelle se déclenche une réaction nucléaire. Appliquée à des activités de recherche, cette métaphore consiste à dire qu'il faut une densité suffisante de chercheurs dans une institution, une métropole ou une région pour que la qualité de la recherche soit bonne, les chercheurs étant censés avoir besoin de nombreux collègues à proximité pour échanger des idées et être stimulés dans leur travail. Toutefois, s'il est possible en physique de calculer avec précision la masse critique, ce n'est nullement le cas dans les activités sociales, celles de recherche comme les autres. Quelques tentatives ont été effectuées pour établir un lien entre le nombre de chercheurs rassemblés dans une même ville ou région et le nombre moyen d'articles publiés par chercheur¹⁸. Elles n'ont pas pu établir ce lien et tout semble indiquer que la masse critique en matière de recherche n'est rien d'autre qu'une idée reçue, sans fondement empirique.

Le nombre de publications d'une ville ou d'une région est en général quasiment une fonction linéaire du nombre de chercheurs, lequel résulte des évolutions de l'enseignement supérieur et des politiques conduites à l'échelle nationale ou locale. Ainsi dans le cas français, la concentration des chercheurs dans la région parisienne ne s'explique nullement par le climat du bassin parisien, les bienfaits de l'urbanisme haussmannien ou les politiques actuelles du Conseil Régional d'Île de France, mais par une très longue histoire du système scientifique d'un pays très centralisé, où les gouvernements ont choisi de concentrer les moyens dans la capitale depuis les grandes écoles du XVIIIe siècle, l'Université Impériale de 1808 ou le CNRS de 1939.

Dans le discours du directeur de l'AERES, la "masse critique" semble être fixée à 500 chercheurs relevant de laboratoires classés A+, mais il ne précise nulle part comment ce seuil a été calculé et à quelle équation il correspond. Le rapport de l'AERES que le directeur présente¹⁹ est une simple compilation par académie des rapports de l'Agence qui ne comporte

¹⁸ Voir par exemple l'article suivant : Bonnacorsi A. et Daraio C., 2005, "Exploring size and agglomeration effects on public research productivity", *Scientometrics*, Vol. 63, n°1, pp.87-120.

¹⁹ Voir <http://www.aeres-evaluation.fr/index.php/Actualites/Actualites-de-l-agence/AERES-2010-Analyses-regionales-des-evaluations-realisees-entre-2007-et-2010>.

aucune analyse spatiale qui mettrait en rapport de façon systématique des éléments d'évaluation avec des caractéristiques locales. L'expression "masse critique" y apparaît 9 fois dans les contextes les plus divers pour évoquer la taille d'écoles doctorales, d'Universités, d'UFR, de laboratoires, etc. Il y a tout lieu de penser que le seuil de masse critique évoqué par son directeur dans l'article cité est parfaitement arbitraire et ne correspond à rien d'autre qu'à l'imaginaire de celui qui le mentionne. (...)"

Annexe 2

Interprétation des notations de l'échelle de notation multicritères

Texte produit par l'AERES, 5 mai 2011

Cette grille explicite l'*interprétation* des notations A+, A, B, C et des niveaux de **performance** requis pour accéder à ces catégories. Les critères s'appliquent à la production, l'impact, la stratégie ou le projet de l'unité de recherche au cours de la période évaluée, typiquement les quatre dernières années. Ils tiennent compte des missions imparties à chaque unité.

Production

A+ : leadership international

La production de cette unité est au meilleur niveau international. C'est une référence incontournable dans son domaine. Plusieurs de ses productions ont été distinguées par des invitations à des conférences majeures, des prix, une valorisation significative.

A : production visible internationalement

La production de cette unité est excellente dans le paysage national. Elle est bien visible au niveau international sans atteindre les positions de leader des unités A+. Bien inscrite dans des actions internationales, elle a apporté plusieurs contributions très importantes.

B : importance nationale

La production de cette unité est solide. Elle contribue significativement à la production de connaissances mais sa visibilité internationale reste limitée.

C : la production de cette unité doit être notablement améliorée pour atteindre la qualité attendue d'unités disposant de moyens et de missions analogues.

Attractivité, Impact socio-culturo-économique

A+ : visibilité internationale

L'unité est attractive au niveau international. Elle attire des chercheurs étrangers de haut niveau et accueille de manière très significative des docs et postdocs avec leur financement, en provenance de l'Europe de l'Ouest, d'Asie, d'Amérique du Nord (alt. **Des pays développés**). Elle est leader (ou PI) dans des programmes internationaux. Ses actions ont un retentissement international.

A : visibilité nationale

L'unité attire des chercheurs français de bon niveau et accueille des docs et post docs de la France entière. Elle participe à des actions européennes et est leader d'actions nationales (GDR, réseaux,...). Ses actions ont un retentissement national.

B : visibilité limitée

L'unité a un impact essentiellement régional. Elle participe à des réseaux nationaux mais pas en position de leader. Son recrutement de permanents et non permanents (docs et postdocs) est très majoritairement **d'origine locale**.

C : Impact insuffisant

L'impact et la visibilité de cette unité restent **limités au niveau local**.

Vie de l'unité

A+ : *excellente tant dans l'animation et la gouvernance de l'unité qu'en ce qui concerne la promotion/émergence de jeunes équipes ou la prise maitrisée de risques. Tous les aspects de la vie de l'unité (séminaires, site web, stratégie d'affectation des moyens humains, techniques et financiers...) sont bien pris en compte.*

A : *très bonne, l'unité a bien mis en œuvre la grande majorité des éléments mentionnés en A+ mais certains d'entre eux sont insuffisamment déployés.*

B : *L'unité devrait à améliorer notablement sa pris en compte, dans le champ du critère, de plusieurs domaines qui sont inexistantes ou mal adaptés. L'unité doit sérieusement développer ses outils de vie collective.*

C : *des problèmes importants doivent être résolus pour permettre à cette unité de se développer et de mettre en œuvre des projets **compétitifs**.*

Stratégie et projet scientifiques

A+ : *excellent projet, novateur, présentant un bon dosage de prise de risque et de mise en œuvre des compétences et originalités de l'équipe/unité.*

A : *très bon projet mais manquant un peu d'ambition, de prise de risque ou ayant fait l'objet d'enjeux moins importants que la catégorie A+.*

B : *projet pouvant présenter un intérêt au niveau national mais insuffisamment construit ou manquant d'ambition ou dont la faisabilité et/ou l'intérêt fait l'objet de critiques de la part du comité d'experts. Le projet devrait être retravaillé avant mis en œuvre.*

C : *projet manquant d'ambition ou dont la faisabilité ou l'intérêt fait l'objet de critiques importantes de la part du comité d'experts.*

Note globale

La note globale intègre les appréciations des experts sur les différents critères ; **elle accorde une place prépondérante à la production de l'unité et tient compte de manière significative de son impact.** Elle met ces éléments en lumière des conditions dans lesquelles la recherche a été effectuée, c'est-à-dire des objectifs et missions impartis à l'unité ainsi que des moyens qui lui ont été attribués ou qu'elle a réussi à mobiliser.

A+ : *unité de très haut niveau, sa visibilité internationale est incontestable, son impact majeur.*

A : *unité de haute qualité à la production remarquable et constituant un élément important du paysage français de la recherche.*

B : *bonne unité mais dont la qualité doit être améliorée dans les domaines précisés par la note multicritérisée et détaillés dans le rapport.*

C : *unité dont la qualité devrait être améliorée de manière importante pour accéder au niveau d'excellence attendu. Des changements majeurs sont attendus dans un ou plusieurs domaines précisés dans le rapport.*

Une grosse unité peut n'être visible internationalement que grâce à une partie restreinte de son effectif. Dans ce cas, elle ne pourra être classée A+.

Annexe 3

L'évaluation contre la science ; productivité académique vs contribution scientifique : le cas de l'économie

Serge Walery (Aix-Marseille Université)
Mai 2011

-Extraits-

Nota : toutes les notes de bas de page ainsi que la bibliographie du texte ont été supprimés.

Certains passages ont été soulignés par nous pour attirer l'attention du lecteur et faciliter une lecture rapide.

Le texte complet est publié dans la livraison de juillet 2011 de la revue *Économie Politique*.

Depuis quelques années, les conditions et modalités d'évaluation de la recherche scientifique ont connu de multiples et profondes transformations. L'harmonisation européenne, la multiplication des comparaisons et classements internationaux, le développement des techniques et des indicateurs bibliométriques, l'influence croissante des listes classant les revues, la systématisation des financements sur projets et/ou appels d'offres, la création en France de l'ANR puis de l'AERES... sont autant d'éléments qui ont participé à ce bouleversement du paysage de l'évaluation de la recherche.

Ces différents éléments "font" système ou régime et (ont un impact), dans le cadre de ce régime, sur le travail des chercheurs, les conditions de production de leurs travaux, et donc, la dynamique de la connaissance scientifique. Ce régime d'évaluation, qu'on qualifiera de "quantitatif", incite au conformisme des chercheurs, et génère une standardisation du travail de recherche conduisant à l'asphyxie progressive de la dynamique de la connaissance scientifique. On remarquera, qu'en économie, cette tendance est sensiblement renforcée par la dérive instrumentaliste de la discipline.

Sur fond de compétition internationale et d'harmonisation européenne, les récentes réformes institutionnelles (AERES, ANR, réforme du CNU...) et le poids croissant des financements (européens, nationaux ou locaux) sur projets, s'inscrivent clairement dans cette perspective.

Du point de vue des managers de la recherche, intimement convaincus de la "flânerie" des chercheurs comme Taylor l'était de celle des ouvriers, le développement et le renforcement des processus d'évaluation visent, par le renforcement de la compétition académique, statutaire et économique, à l'amélioration de la productivité (intensification) du travail des chercheurs et des laboratoires. Or, comme pour toute autre activité, l'affirmation de la logique manageriale dans le monde de la recherche s'accompagne nécessairement d'une obsession quantitativiste.

La montée en puissance du Régime d'évaluation quantitative (REQ) est également le fruit de la réalité matérielle à laquelle se trouvent confrontés les évaluateurs. Comme le souligne un récent rapport de l'Académie des Sciences, les modalités traditionnelles d'évaluation sont dévoreuses de temps.

À tous les échelons, les instances d'évaluation sont concrètement dépassées par la masse des dossiers à évaluer. Les membres de ces institutions, contraints de rationaliser leur travail

d'évaluation, sont naturellement conduits à recourir à des indicateurs quantitatifs. Ceux-ci accélèrent considérablement la prise de décision, en réduisant le temps nécessaire au traitement de chaque dossier, en "évacuant" la complexité des appréciations des rapporteurs, en simplifiant les hiérarchisations et classements, et surtout, en limitant, voire en supprimant, les débats et les controverses. Comme dans bien d'autres domaines, compter permet de ne pas débattre, et se révèle infiniment plus simple et plus rapide.

Il convient enfin de souligner le rôle des chercheurs eux-mêmes, dans le renforcement de l'évaluation quantitative de la recherche. D'abord, parce qu'en vertu du principe traditionnel d'évaluation par les pairs, ils sont souvent évaluateurs. Mais aussi, parce qu'ils ont globalement (du moins en économie), joué sincèrement le "jeu" qui leur était proposé. Si l'esprit de compétition, dont on ne voit guère pourquoi il serait moins développé chez les chercheurs que chez les autres individus, fut longtemps encadré par les normes et conventions traditionnelles du monde de la recherche, la montée en puissance de la logique managériale a conduit à le libérer et à l'encourager. Les chercheurs ont ainsi activement contribué au renforcement de la compétition symbolique et statutaire, ainsi qu'à l'émergence et à la montée en puissance d'une compétition économique et bibliométrique de plus en plus exacerbée.

L'esprit de compétition et la soif de classement des chercheurs ont ainsi convergé avec la logique managériale et le poids des contraintes matérielles, pour pousser à la généralisation et à la systématisation de l'évaluation quantitative de la recherche. Au sein des sciences sociales, l'économie, par nature familière du quantitatif, et depuis longtemps travaillée par un fantasme de rapprochement avec les sciences "dures", s'est logiquement retrouvée à la pointe du mouvement.

Du point de vue du manager de la recherche, de l'évaluateur ou du chercheur, le recours systématique, voire exclusif, aux indicateurs quantitatifs et aux classements qui en découlent, présente un quadruple intérêt : simplifier et rationaliser le travail d'évaluation, circonscrire les débats et controverses, donner une apparence d'objectivité aux classements obtenus, et "cardinaliser" les objectifs de la compétition des chercheurs.

C'est ainsi que, depuis une douzaine d'années, le CNU-5 s'est doté, pour la qualification des MCF, d'une "grille d'évaluation" comportant dix indicateurs, dont sept purement quantitatifs, les autres l'étant potentiellement. Ou que la liste recensant et classant (à l'origine, par un nombre d'étoiles) les revues d'économie et de gestion, établie par la section 37 du CNRS, a généré, dès sa première publication en juillet 2004, une véritable "course aux étoiles". Objectif désormais prioritaire de bien des chercheurs, qui le font fréquemment figurer sur leur curriculum vitae, le "nombre d'étoiles CNRS" est, du point de vue de l'évaluateur ou du manager, un outil de classement particulièrement simple et pratique. Même si le CNU-5 terminant actuellement son mandat a repoussé de justesse l'utilisation explicite d'un cardinal de ce type pour les promotions des enseignants-chercheurs, la liste CNRS-37 n'en demeure pas moins une référence implicite, et elle a sert de base à l'établissement de "scores de publication" par le jury du concours d'agrégation, depuis 2008. Elle est aussi devenue la référence pour l'évaluation des économistes et de leurs laboratoires par l'AERES, et joue un rôle déterminant dans l'attribution des financements, voire, pour dans un nombre croissant d'institutions, dans la rémunération individuelle des chercheurs (sous forme de primes).

Le fait que les concepteurs et initiateurs de cette liste aient été très largement débordés par son succès et par les usages qui en ont été faits, illustre à la fois, la fascination qu'opèrent les indicateurs quantitatifs sur les chercheurs, évaluateurs et managers, et la puissance que ceux-ci confèrent, par la convergence de leurs comportements, à la dynamique du REQ.

De l'évaluation par les pairs à l'évaluation "par les revues et les listes"

Concrètement, l'affirmation du REQ s'est traduite par de profondes transformations des instruments, du processus, et du travail d'évaluation. Ces transformations se sont articulées avec un "déplacement" de la substance du pouvoir d'évaluation.

Au fur et à mesure que s'affirmait le REQ, les indicateurs ont fait l'objet de multiples perfectionnements et innovations, tel le h-index devenu, en quelques années, l'alpha et l'oméga de nombre de chercheurs, en premier lieu aux États-Unis, et plus vite en sciences "dures" qu'en sciences humaines et sociales. Il convient cependant de rappeler que les indicateurs, aussi perfectionnés et raffinés soient-ils, reposent tous, in fine, sur des données dont le recueil et la construction sont, dans bien des cas, pour le moins approximatifs. C'est ainsi, que les deux grilles remplies (par chacun des deux rapporteurs) pour chaque candidat à la qualification des MCF par le CNU-5, diffèrent fréquemment, en particulier sur les indicateurs quantitatifs, alors même que ces indicateurs sont élémentaires.

"Il est important d'insister sur le fait que nombre d'analyses bibliométriques sont erronées en raison d'un mauvais recueil des données, d'erreurs dans le calcul des indicateurs, ou d'une interprétation inexacte." [Académie des Sciences 2011]

C'est pourquoi on ne s'attardera pas sur la critique technique des indicateurs, certes nécessaire, mais qui, bien souvent, tend à éclipser la réflexion sur la logique et le fonctionnement concret du REQ ; la critique des indicateurs, aussi utile soit-elle, ne saurait se substituer à celle de l'évaluation par indicateurs.

La montée en puissance du REQ s'est traduite par une profonde évolution du travail concret des instances d'évaluation. Le cadre traditionnel de collectifs de scientifiques émettant, après lecture et discussion, des avis (par nature discutables) sur les travaux qui leur sont soumis, se réduit de plus en plus aux apparences. Désormais, dans la plupart des cas, leur tâche effective se limite à opérer la synthèse d'un certain nombre d'indicateurs.

Les conditions matérielles (délais, nombre de candidats, nombre et compétence disciplinaire des membres des Comités de sélection...) poussent souvent, au moins pour certaines étapes à procéder sans lecture des travaux. Les "bonnes intentions", telles celles exprimées par l'Académie des Sciences, ne pèsent guère face à la réalité matérielle.

L'AERES et l'ANR, par leur composition et leurs critères d'évaluation, constituent, à ce jour, l'aboutissement de cette logique : une évaluation sans lecture ni appréciation des travaux. Dès lors, le travail d'évaluation (la synthèse d'indicateurs) ne nécessitant aucune compétence sur le fond, le confier à des chercheurs de la discipline constituerait un évident gaspillage de ressources. D'ailleurs, quel membre du CNU-5, noyé dans le dossier d'un candidat, et cherchant désespérément à en extraire les données nécessaires à l'établissement de la grille d'évaluation, ne s'est jamais dit que ce travail ardu ne nécessite pas la moindre compétence en économie ?

Les différentes instances officielles d'évaluation des chercheurs et de la recherche tendent ainsi à n'être que de simples "chambres" de synthèse d'indicateurs. De ce fait, et malgré les apparences institutionnelles, la substance du pouvoir d'évaluation leur échappe. En simplifiant à l'extrême, pour un économiste français, la substance du pouvoir d'évaluation est désormais entre les mains des comités éditoriaux des revues, à hauteur du classement de la revue dans la liste établie par la section 37 du CNRS. Du score qu'il atteindra dans la "course aux étoiles", dépendront sa carrière (à travers le CNU, le CNRS, l'AERES...), et sa capacité à obtenir des financements (Universités autonomes, ANR, fonds européens...).

Finalement, le REQ, qui doit sa puissance à la conjonction des comportements de tous les acteurs concernés, vise à maximiser la productivité académique des chercheurs, par la

systematisation de la compétition, et l'établissement de classements explicites ou implicites reposant sur des indicateurs quantitatifs. Dans le cadre d'un tel régime, la substance du pouvoir d'évaluation revient, in fine, aux revues et aux listes.

Incitations, biais, contraintes

Du point de vue des chercheurs, le REQ agit comme un système d'incitations visant à les encourager à participer à la compétition, en particulier bibliométrique et économique. Certes, parmi les chercheurs, comme au sein de tout groupe d'individus, il en existe de plus ou moins perméables, ou de plus ou moins rétifs, aux incitations. Mais globalement, au fur et à mesure de la montée en puissance du REQ, les chercheurs ont naturellement été amenés à faire évoluer leurs comportements (objectifs, travail concret...), et donc le produit de leur travail. C'est pourquoi on tentera de repérer les particularités (biais, contraintes) comportementales générées par le REQ, et de montrer comment elles influent sur la nature et le contenu des recherches menées. On s'attachera, en particulier, à montrer comment le REQ pèse sur les capacités d'innovation et de renouvellement des disciplines, et spécifiquement de l'économie.

Idéalement, le "programme" du chercheur se limite à un objectif unique : consacrer toutes les ressources dont il dispose, à faire progresser sa discipline. Bien entendu, la réalité n'a jamais été aussi simple ; l'intérêt individuel des chercheurs et les relations concurrentielles qu'ils entretiennent n'ont pas attendu l'époque contemporaine pour interférer avec cet idéal. Cependant, avec le REQ, il ne s'agit plus d'interférences, mais de substitution ; désormais, les progrès de la discipline ne font plus partie du programme individuel du chercheur. Qu'il s'agisse de sa carrière, de sa reconnaissance académique, de sa capacité à obtenir des contrats de recherche, de la prospérité et du devenir de son laboratoire, et même de son Université, l'objectif exclusif de l'individu-chercheur rationnel est désormais de maximiser "ses" indicateurs. Il serait même possible de ramener le programme du chercheur à la maximisation d'un indicateur synthétique unique, tel que le h-index ou le nombre d'étoiles CNRS. Dans la suite du propos, on considèrera cependant que le chercheur vise à maximiser le nombre et le rang de ses publications, et, à un moindre degré, le nombre de citations dont ses publications font l'objet.

Tout pousse le chercheur à se focaliser sur le nombre de ses publications. C'est de cette obsession que résultent les pratiques de sur-publication.

Il est ainsi logique de tronçonner les résultats d'une recherche en autant d'éléments publiables que possible, quitte à en affaiblir la cohérence d'ensemble ; c'est ainsi que de plus en plus de thèses en économie sont, implicitement ou explicitement, des recueils d'articles. De même, il apparaît rationnel, d'établir successivement, plusieurs "versions" ou "évolutions" d'un contenu similaire, ne différant que par des adaptations et modifications aussi mineures ou formelles que possible, puis, au fur et à mesure de leur rédaction, de les soumettre à des revues différentes. Plus généralement, le REQ pousse le chercheur rationnel à consacrer toute son énergie à la production accélérée de "nouveauautés éditoriales", aussi marginales et insignifiantes soient-elles du point de vue des progrès de sa discipline.

D'autre part, la course à la publication pousse au renforcement de la spécialisation et de son inertie ; le chercheur est, de fait, incité à se spécialiser dans un créneau précis, aussi porteur que possible, puis à exploiter le filon tant qu'il offre des opportunités de publication, puisque, sauf cas particulier, un changement de spécialisation se traduirait par une réduction, au moins temporaire, de son rythme de publication (coûts de transfert).

Apprécier l'ampleur de la sur-publication générée par le REQ est évidemment très difficile. Un rapport commandé par le Ministère de l'Enseignement Supérieur et de la Recherche en fournit un ordre de grandeur, pour l'économie. Il constate que, depuis le milieu des années 1990, le nombre annuel d'articles "Econlit" publiés par des chercheurs français a

approximativement triplé, mais que la “qualité moyenne” des publications n’a, elle, pas progressé. Parallèlement, les auteurs soulignent que l’accroissement du rythme de publication est essentiellement le fait des “entrants”, c’est-à-dire, des générations de chercheurs ayant été, dès le début de leurs carrières, confrontées à la montée en puissance du REQ. On retrouve logiquement cette distorsion générationnelle dans les résultats des évaluations fondées sur des indicateurs bibliométriques ; lors de la session du CNU-5 de mai 2009, parmi les huit promus au premier échelon de la classe exceptionnelle des Professeurs, sept avaient moins de cinquante ans.

Du point de vue des progrès de la connaissance scientifique, la sur-publication, et les innombrables redondances qu’elle entraîne, sont l’un des principaux gaspillages de ressources générés par le REQ : gaspillage lié à la spécialisation et à son inertie ; mais aussi, pour les lecteurs, de plus en plus submergés par une marée de publications, dont l’apport unitaire est souvent très réduit ; et surtout pour les auteurs, par le temps et l’énergie consacrés à multiplier les articles et les applications, plutôt qu’à essayer de progresser sur le cœur de leur démarche. Plus généralement, la pression induite par le REQ conduit le chercheur rationnel à consacrer un temps considérable à communiquer sur son travail, que ce soit par des participations à des colloques, indispensables pour accroître ses probabilités de publication et de citation, ou par la rédaction d’articles aussi nombreux que possible. Au fond, le chercheur rationnel consacre infiniment plus de temps et d’énergie à faire en sorte de publier, qu’à chercher. Il convient de souligner les effets de ce biais sur les jeunes chercheurs, auxquels les impératifs de publication interdisent désormais de prendre le temps nécessaire pour se constituer une culture générale de leur discipline, pour construire peu à peu leur propre démarche, leur propre cadre de pensée.

Si elle accroît la productivité académique telle qu’elle est aujourd’hui mesurée, la sur-publication pèse ainsi négativement sur la productivité scientifique ; en outre, en favorisant l’inertie des spécialisations, elle réduit les capacités de renouvellement, et contribue ainsi à asphyxier ou stériliser la dynamique de la connaissance scientifique.

Alors que l’ouvrage avait longtemps été la forme privilégiée de diffusion des résultats de la recherche en sciences sociales, la montée en puissance du REQ a conduit, en particulier en économie, à s’aligner sur les pratiques des sciences “dures”, en faisant de l’article (dans une revue à comité de lecture) la forme quasi-exclusive de la production académique.

En sciences sociales, la contrainte concrète la plus immédiate que fait peser le format article, est la limitation stricte de la “taille” des textes produits. Si chaque revue est libre de définir la limite qu’elle souhaite, il existe de fait une norme courante qui, en économie, est de l’ordre de 50 à 90 000 signes. Quels que soient ses raisons et ses éventuels effets positifs, cette contrainte conditionne largement le travail des chercheurs.

Tout d’abord, le format article confronte le chercheur à un dilemme : s’il souhaite rendre compte précisément et rigoureusement des différentes étapes de son raisonnement, il est condamné à se limiter à un objet étroitement délimité, à un champ réduit, voire ponctuel ; s’il souhaite aborder un champ élargi, il est condamné à un propos très synthétique, tendant à gommer les nuances, les détails et les précautions qui caractérisent la démarche scientifique. En économie, où la première option a clairement été privilégiée par les revues et les chercheurs, la prépondérance du format article a ainsi joué un rôle important dans le renforcement de la spécialisation des chercheurs et du morcellement de la discipline.

Parallèlement, le format article, en incitant les économistes à privilégier les outils et méthodes qui s’y adaptent le plus aisément, stimule certains domaines de la pensée économique, en déstabilise d’autres, et, par conséquent, influe directement sur la dynamique et la structuration de la discipline. Il a ainsi contribué au recours de plus en plus systématique à la formalisation

mathématique et aux tests économétriques, et, plus récemment, au succès de l'économie expérimentale. À l'inverse, alors que d'éminents économistes, tel Maurice Allais, dénoncent une "spécialisation outrancière", plaident pour un effort général de synthèse, et regrettent le manque de culture historique des jeunes économistes, le format article favorise une histoire de la pensée économique de plus en plus spécialisée, voire "ponctuelle", et philologique ; quant à l'économie historique, la quasi-impossibilité de présenter un travail de synthèse dans le format article entrave considérablement son intérêt pour les non-spécialistes, et donc son rayonnement.

On peut enfin remarquer que le format article est porteur d'une vision très déformée, voire artificielle, du travail scientifique. Les interrogations qui demeurent, les doutes, les fausses pistes et les errances de la démarche scientifique sont, en effet, de plus en plus systématiquement passés sous silence. D'une part, les revues ne publient pas les fausses pistes et les errances, d'autre part, le format article ne laisse guère l'opportunité d'en parler. Sans compter que la plupart des chercheurs sont convaincus, à tort ou à raison, qu'évoquer les méandres de la démarche scientifique réduit la probabilité de publication.

Les chercheurs français sont confrontés à une autre contrainte formelle lourde ; celle de la langue, et plus précisément de l'anglais. Certes les effets de cette contrainte linguistique paraissent, a priori, limités pour les sciences "dures" ; la présentation d'une démonstration mathématique ne dépend que très marginalement de la langue utilisée, et n'exige pas de son auteur une parfaite maîtrise de cette langue. Cependant, en sciences humaines et sociales, où la langue écrite demeure le principal véhicule de la pensée, rendre compte rigoureusement de phénomènes complexes exige une maîtrise très précise de la langue utilisée, et le passage d'une langue à l'autre soulève fréquemment des problèmes de fond, qui donnent lieu à d'innombrables débats de traduction.

On peut également souligner la puissance de cette dérive instrumentaliste dans l'enseignement de l'économie.

En économie, la contrainte linguistique est particulièrement forte ; l'immense majorité des revues référencées sont anglophones. Surtout, la prépondérance des revues anglophones est d'autant plus forte qu'on s'élève dans la hiérarchie des revues, telle qu'elle est définie par les listes et les classements. Ainsi les six revues économiques³⁰ classées 1* (rang le plus "élevé") dans la liste CNRS-37 sont anglophones. Quant aux 58 (sur 705) revues francophones classées, 40 d'entre elles sont en catégorie 4, la plus faible. On peut aussi remarquer que la seule revue francophone classée en catégorie 1 (*Annales, Histoire, Sciences sociales*) ne publie, paradoxalement, que très peu d'articles d'économie.

Comme le format article, la contrainte linguistique ne pèse pas de la même façon sur les différents domaines de l'économie, et influe donc sur la dynamique et la structuration de la discipline.

Normalisation, conformisme, standardisation

Ces biais et contraintes encadrant étroitement le travail concret des chercheurs, s'inscrivent dans un processus général de normalisation des comportements induit par le REQ. Dans ce processus, les revues, auxquelles a, de fait, été délégué le pouvoir d'évaluation, sont le "lieu" de production des normes et des modes, que le conformisme et les incitations du régime d'évaluation poussent les chercheurs à incorporer. Il en résulte une tendance à la standardisation de la production scientifique, qui affaiblit les capacités d'innovation et de renouvellement des disciplines, et, plus largement, la dynamique de la connaissance scientifique.

En économie, ce processus prend des formes originales. Par rapport aux autres sciences sociales, la normalisation instrumentale y joue un tel rôle, qu'on peut parler de dérive

instrumentaliste ; insensiblement, dans la démarche scientifique de bon nombre d'économistes, "l'outil est devenu l'objet".

Dans un milieu professionnel par nature sujet au conformisme et aux jeux de réseaux, ne serait-ce que par le principe de cooptation, le pouvoir des revues dans le REQ se déploie à travers le réseau de chercheurs qui gravite autour de chaque revue. Le centre de gravité de ce réseau est constitué par les membres du board et les referees. Bien sûr, en tant que chercheur, chacun d'entre eux, par son appartenance à d'autres réseaux professionnels, dispose, de fait, d'un réseau de relations, plus ou moins proches, qui lui est propre. C'est l'ensemble de ces relations avec le centre de gravité qui constitue le réseau d'une revue.

Or, appartenir au réseau gravitant autour d'une revue, directement ou par le réseau personnel d'un membre du board ou d'un referee, constitue, pour un chercheur, un avantage concurrentiel significatif pour publier dans cette revue, ne serait-ce que par l'accès privilégié à l'information sur les normes (explicites et implicites) de publication propres à la revue, ou aux éventuels colloques au sein desquels il est possible de faire connaître et de "roder" son article auprès des membres du réseau.

Se constitue un enchevêtrement hiérarchisé de réseaux de revues, qui constitue le principal vecteur de diffusion, d'incorporation et de reproduction, des normes et des modes. C'est par son intermédiaire que s'opèrent la normalisation et la standardisation du travail des chercheurs.

En déléguant la substance du pouvoir d'évaluation aux revues, le REQ leur confie, du même coup, le pouvoir de définir les normes et les modes propres à chaque discipline. Chaque revue, par son positionnement (théorique, méthodologique...), par son domaine de référence, et par les membres de son réseau, est porteuse d'une certaine vision de la discipline, de ses paradigmes, de ses thématiques, de ses méthodes. À une échelle plus globale, c'est à travers l'enchevêtrement hiérarchisé de réseaux, au sein duquel les revues les mieux classées sont logiquement les plus influentes, qu'est générée la vision dominante et normalisatrice de la discipline, vision bien sûr évolutive. Ce sont ainsi les revues, et en particulier les mieux classées, qui orientent et canalisent les évolutions et le devenir de chaque discipline. En économie, où la suprématie des revues anglo-saxonnes, et en particulier américaines, frise le monopole, le REQ a considérablement renforcé l'emprise, déjà forte, des normes et de la vision "anglo-saxonnes" de la discipline. Les économistes contemporains sont ainsi, beaucoup plus que les générations précédentes, touchés par une normalisation de la discipline à l'échelle mondiale.

Ce processus de normalisation est particulièrement puissant, car, loin de tenter d'y résister, le chercheur rationnel, attentif aux incitations induites par le REQ, consacrera, au contraire, d'importantes ressources à l'incorporation des normes produites par les revues. Pour accroître sa probabilité de publication dans une revue, il s'efforcera naturellement d'appréhender et de faire siennes la vision de la discipline et les normes thématiques, méthodologiques, instrumentales et rédactionnelles, propres à ladite revue. Et pour multiplier le nombre et améliorer le rang de ses publications, il sera logiquement conduit à tout faire, pour coller au plus près de la vision dominante de la discipline et de ses évolutions.

Certes, rien n'interdit à des chercheurs ne se reconnaissant pas dans la vision dominante, d'organiser séminaires et colloques, et de publier des ouvrages, individuels ou collectifs ; cependant, au moins en économie, ces formes de production scientifique sont de plus en plus ignorées par les instances d'évaluation. De même, rien n'interdit d'éditer des revues porteuses d'une vision hétérodoxe de la discipline, mais leur référencement et leur classement dans les listes nuiront inévitablement à la carrière des auteurs y publiant leurs travaux. Le REQ tend ainsi à exclure les chercheurs prétendant échapper à la vision dominante, en les condamnant à une carrière de second rang.

Le REQ, conjugué à la systématisation des financements sur projets, rend ainsi le conformisme rationnel pour le chercheur. Logiquement, celui-ci ne produira que ce qui a une forte probabilité de publication, et ne se lancera que dans des projets dont les résultats sont prévisibles, programmables, rapides et publiables. Il en résulte un comportement général d'autolimitation, visant à se conformer aux normes et impératifs des revues, et générant une tendance lourde à la standardisation des ambitions, des perspectives, des paradigmes, des thématiques, des méthodes et outils, des pratiques rédactionnelles...

Bien entendu, d'une discipline à l'autre, la pondération des différents "axes" de la standardisation est variable. Ainsi l'économie contemporaine est-elle très ouverte en matière de thématiques ; de Gary Becker (Nobel 1992) à Steven Levitt (John Bates Clark Medal 2003), des économistes reconnus se sont penchés sur des questions tellement variées et apparemment peu économiques, qu'on parle aujourd'hui d'"économie saugrenue" [Levitt, Dubner 2005]. De même, à l'heure du défi posé par les répercussions de la crise financière et de la contrainte écologique, deux des six revues économiques les plus prestigieuses viennent de publier des articles sur des sujets tels que, les liens existant entre la violence conjugale et les résultats des matchs de football américain, ou l'avantage dont disposerait l'équipe tirant en premier, lors des séances de tirs au but qui concluent les matchs de football, en cas d'égalité après prolongations.³⁸ À l'inverse, l'économie contemporaine souffre d'une très forte standardisation paradigmatique : la représentation walrassienne de l'économie comme ensemble de marchés interdépendants tendant vers l'équilibre. Elle est en outre, du fait de sa dérive instrumentaliste, particulièrement soumise aux modes instrumentales ; à la vogue de l'économie quantitative de la fin du 20^{ème} siècle, s'est ainsi ajoutée, au début du 21^{ème}, celle de l'économie expérimentale, voire de la neuro-économie.

Cette standardisation des comportements et du travail des chercheurs est évidemment nuisible à leur créativité, à leur imagination, à leurs ambitions, à leur esprit critique, à leur goût du risque... En incitant les chercheurs à consacrer leur énergie et leur intelligence à l'incorporation des normes et à la satisfaction des critères d'évaluation, elle génère d'importants gaspillages de ressources, du point de vue des progrès de la connaissance scientifique. Le développement de la connaissance scientifique, les capacités d'innovation des chercheurs et le renouvellement des disciplines, sont ainsi étroitement canalisés et bornés.

Appréciation vs évaluation

Le REQ, par les biais comportementaux qu'il génère, par le pouvoir normalisateur qu'il délègue aux revues et aux listes, et en incitant le chercheur rationnel au conformisme, débouche sur une standardisation du travail des chercheurs, et donc du produit de leur travail. Visant à améliorer la productivité académique, il tend concrètement à asphyxier la dynamique de la connaissance scientifique. En paraphrasant Nietzsche, on pourrait dire qu'à trop vouloir accroître la productivité académique, le REQ affaiblit la productivité scientifique.

De façon qui n'est paradoxale qu'en apparence, le REQ se révèle ainsi être un obstacle important au renouvellement de la pensée économique, dont la crise financière récente et les préoccupations écologiques ont cruellement mis en évidence la nécessité. Promouvoir un tel renouvellement exige donc, parmi bien d'autres choses, d'enrayer la dérive quantitativiste qui affecte aujourd'hui le régime d'évaluation.

En ce qui concerne les chercheurs, dissiper l'illusion quantitativiste ne devrait pas poser de problèmes insurmontables ; ils savent tous que les progrès scientifiques ne sont pas affaire de nombre de publications, ni même de citations.

Les chercheurs savent tous que, bien souvent, les progrès les plus décisifs sont le fruit de "recherches orphelines", que le régime d'évaluation quantitative, par nature, ignore.

Ces constats n'empêchent pourtant pas la plupart des économistes de se satisfaire du REQ, ou de considérer comme perdue d'avance toute tentative de remise en cause. Quant aux rares résistances qui se font jour, elles tendent malheureusement à se cantonner à la défense des modalités "traditionnelles" d'évaluation, lesquelles n'étaient, faut-il le rappeler, ni exemptes de tout biais, ni irréprochables en matière d'incitations au conformisme

Il ne saurait évidemment être question de proposer ici un régime d'évaluation "clés en mains", alternatif au REQ. Tout au plus peut-on esquisser quelques pistes de réflexion, s'appuyant, certes, sur la réaffirmation de certains principes traditionnels, mais s'efforçant de ne pas trop succomber à l'illusion du retour à un âge d'or qui n'a jamais existé. Ces esquisses s'inscrivent dans la perspective tracée par l'Association Française d'Économie Politique, dans la mesure où elles s'appuient sur la conviction que la promotion systématique du pluralisme et de la diversité, en particulier sur les plans méthodologique et instrumental, est un des moyens privilégiés de lutte contre la sclérose et l'infertilité de la pensée économique contemporaine.

Enrayer la dérive quantitativiste du régime d'évaluation suppose, en premier lieu, que la substance du pouvoir d'évaluation ne soit plus accaparée par les revues et les listes, mais reprise par les instances qui en sont officiellement dotées. Or, une telle ré-appropriation ne peut être envisagée que si les instances prévues à cet effet en reviennent à une évaluation sur travaux (et non sur indicateurs). La dérive quantitativiste ne saurait être combattue sans que soit systématiquement réaffirmé le principe selon lequel toute évaluation sans lecture et appréciation des travaux devrait être proscrite.

De ce principe élémentaire découlent, concrètement, de multiples effets ou conséquences. La lecture et l'appréciation des travaux (ou d'une sélection de travaux) permettraient, tout d'abord, de rendre leur importance aux formats autres que l'article, laissant ainsi plus de place à des démarches moins "standard". Parallèlement, récuser toute évaluation sans lecture des travaux imposerait, matériellement, une réduction significative du nombre et de la fréquence des évaluations, laquelle favoriserait la qualité effective de l'évaluation, et limiterait, au moins à court terme, les incitations à la sur-publication. Enfin, l'application de ce principe permettrait, selon les préconisations de l'Académie des Sciences [Académie des Sciences 2011], d'interdire le recours aux indicateurs bibliométriques pour l'évaluation des jeunes chercheurs, et d'en circonscrire étroitement l'usage, pour les autres.

Le principe de lecture et d'appréciation des travaux induit, inévitablement, celui d'évaluation par les pairs. Or, le passé de l'Université française a abondamment montré que l'évaluation par les pairs pouvait être, elle aussi, profondément conformiste et génératrice de biais. Il importe donc de ne pas se contenter d'un simple retour au principe d'évaluation par les pairs, mais d'organiser et de promouvoir systématiquement la diversité et le pluralisme, tant des instances qu'au sein des instances.

Selon la même logique, les modalités de fonctionnement et de délibération des instances d'évaluation devraient être élaborées et adaptées dans le souci permanent de valoriser le pluralisme, de mieux prendre en compte les recherches orphelines et/ou hétérodoxes, et de réfréner le conformisme "naturel" des chercheurs. Parallèlement, la logique d'examen devrait être systématiquement privilégiée, la logique de concours n'intervenant qu'en second temps, et seulement dans les cas où elle se révèle indispensable.

Pour soutenir aujourd'hui les progrès de la connaissance scientifique, et spécifiquement pour œuvrer au renouvellement de l'économie, il convient enfin de mener le combat du vocabulaire, beaucoup moins anodin qu'il n'y paraît. Face aux impasses de l'évaluation quantitative de la productivité académique, il importe de promouvoir, inlassablement, une logique d'appréciation de la contribution scientifique des chercheurs et des laboratoires.

Annexe 4

La loi relative aux libertés et responsabilités des Universités : essai de mise en perspective

Annie Vinokur

Revue de la régulation.

Capitalisme, institutions, pouvoirs

[n°2, Janvier 2008](#)

Opinions – débats

-Extraits-

Nota : toutes les notes de bas de page ainsi que la bibliographie du texte ont été supprimés.

Certains passages ont été soulignés par nous pour attirer l'attention du lecteur et faciliter une lecture rapide.

Parler de *politique* suppose d'inscrire les choix collectifs dans un projet de société publiquement exprimé et justifié ; lui substituer celui de *management* contient implicitement l'idée qu'il n'y a pas de débat possible – ou nécessaire – sur les fins, mais seulement une expertise à développer sur les moyens.

Le discours européen est passé de l'objectif de *lisibilité* interne à celui de *visibilité* mondiale, qui ne peut procéder que de la poursuite individuelle darwinienne de *l'excellence*.

La loi LRU, quant à elle, a pour vocation de donner aux établissements universitaires la capacité d'être de bons opérateurs de la LOLF en les rapprochant du modèle de l'Université entrepreneuriale. Le pouvoir de la direction est renforcé au détriment de celui des personnels et des étudiants dans les conseils, et les domaines où ce pouvoir peut s'exercer sont élargis:

Financement : liberté de recherche de ressources extrabudgétaires dans des activités commerciales et des revenus de fondations

Inputs : maîtrise du patrimoine immobilier, maîtrise de la masse salariale, contrôle accru des recrutements d'enseignants statutaires, embauche facilitée de personnels contractuels (sous réserve de quotas fixés par contrat par le ministère pour chaque établissement), modulation des rémunérations.

Le problème (des Universités) est celui de l'héritage d'un sous-financement. La LRU "répond" à ce problème en prévoyant, au niveau des laboratoires, une substitution aux dotations récurrentes d'un financement par projets piloté par une agence de moyens (ANR) dont les membres sont nommés et non élus, ainsi qu'une évaluation de la production scientifique sur des critères (bibliométrie, brevets...) désormais largement disqualifiés, selon elle, dans la communauté scientifique internationale.

L'offensive laisse à penser que, comme aux États-Unis à partir des années 1980, l'objectif immédiat de la réforme est d'en finir avec un pouvoir du savoir jugé politiquement dangereux et économiquement insuffisamment rentable.

Annexe 5

Du travail à l'action publique : quand les dispositifs d'évaluation prennent le pouvoir

M.-C. Bureau

Cahiers internationaux de sociologie, 2010

-Extraits-

Nota : toutes les notes de bas de page ainsi que la bibliographie du texte ont été supprimés.

Certains passages ont été soulignés par nous pour attirer l'attention du lecteur et faciliter une lecture rapide.

Les mécanismes d'évaluation exercent aujourd'hui une emprise croissante sur nos vies, bien au-delà de la période scolaire rythmée par les notes et les examens. Ils prennent des formes variées mais de plus en plus interdépendantes entre elles. Ainsi, l'évaluation du travail avec ses effets sur la rémunération, les chances de promotion et les pratiques professionnelles, l'évaluation des organisations avec ses conséquences pour le devenir des structures, l'évaluation de l'action publique avec son retentissement sur le cours de la vie politique, tendent à faire système. Ces différentes formes concourent au fonctionnement d'un dispositif global, au sens de Michel Foucault, qui assure le pouvoir de la norme sur la population des individus, au moyen d'un jeu incessant de visibilité, diagnostic et classement

Et pourtant, nous constatons dans ce domaine un désarmement de la critique, ou, plus précisément, à l'exception de quelques dénonciations qui s'en prennent à l'idée même de quantification du social, un paradoxe entre la légitimité persistante du principe d'évaluation et l'illégitimité récurrente des dispositifs d'évaluation tels qu'ils sont effectivement mis en œuvre dans de nombreux domaines de la vie sociale (Demailly, 2006). En conséquence de ce paradoxe, les effets de l'évaluation sont plus souvent indirects que directs : la normalisation des conduites, amplifiée par la dynamique des anticipations, a de fait un impact beaucoup plus fort sur nos existences qu'une myriade de dispositifs fluctuants, segmentés, contestés et régulièrement détournés par les acteurs sociaux.

L'hypothèse que nous défendons ici est la suivante : ce qui rend si difficile la critique de l'évaluation, y compris dans les formes les plus pathogènes qu'il lui arrive de revêtir aujourd'hui, c'est le fait qu'elle allie l'exercice de la rationalité instrumentale, caractéristique de la modernité, avec la poursuite de finalités éthiques peu contestées telles que la justice ou la démocratie.

Un tel constat s'applique aussi bien à la montée irrésistible, à laquelle nous assistons depuis environ un demi-siècle, d'une culture de l'évaluation. En effet, celle-ci procède à l'évidence de l'entreprise moderne de rationalisation, de cette conviction que l'on peut en principe maîtriser toute chose par le calcul, mais elle acquiert surtout ses lettres de noblesse par les objectifs qu'elle se donne : la reconnaissance du travail à sa juste valeur, la transparence démocratique et l'efficacité de l'action publique.

De l'évaluation du travail à celle de l'action publique, les dispositifs font aujourd'hui système : les indicateurs de performance se déclinent aux niveaux individuel et collectif ; les palmarès et les classements orientent les politiques publiques comme les stratégies d'établissements ou les conduites individuelles ; les conventionnements sur projet, tributaires des évaluations, retentissent directement sur les pratiques de recrutement et de rémunération. Ces différents rouages participent d'une même technologie de gouvernement, résumée par la

formule proverbiale du benchmarking : "qui veut s'améliorer doit se mesurer, qui veut être le meilleur doit se comparer".

Le pilotage de l'activité par des indicateurs quantitatifs, dès qu'il est prévu par la LOLF, génère à profusion les effets pervers, parfois aux limites de l'absurdité.

(Évalué), le travail tend (aussi) à s'organiser autour des indicateurs et, si la pression sur les objectifs s'accroît, le salarié peut être amené à sacrifier pour partie la finalité de son métier ainsi que ses propres règles déontologiques.

Si la machinerie de l'évaluation apparaît donc à ce point vulnérable à la critique et si facile à détourner, pourquoi exerce-t-elle une telle emprise sur nos vies ? C'est là qu'il nous faut revenir à la notion de dispositif, tel que Giorgio Agamben (2006) la définit à la suite de Foucault : "J'appelle dispositif tout ce qui a d'une manière ou d'une autre la capacité de capturer, d'orienter, de déterminer, d'intercepter, de modeler, de contrôler et d'assurer les gestes, les conduites, les opinions et les discours des êtres vivants. (...) Le dispositif est donc, avant tout, une machine qui produit des subjectivations et c'est par quoi il est aussi une machine de gouvernement". Là réside le principal pouvoir de la batterie d'indicateurs, d'outils et de protocoles qui constituent la machinerie de l'évaluation : elle transforme nos manières de sentir, de raisonner et de réagir même lorsque les dispositifs eux-mêmes sont peu opérants et peu légitimes. Quelles que soient les critiques raisonnables qui sont adressées à tel ou tel de ses rouages, la machinerie a acquis une autonomie suffisante pour survivre, produire des subjectivations et influencer durablement nos conduites. Par exemple, le fait que les critères de construction du palmarès de Shanghai (classement des Universités au niveau mondial) aient été critiqués de façon très pertinente, n'enlève pas grand-chose à l'impact de celui-ci sur les subjectivités : il s'est d'une certaine façon affranchi des conditions qui ont présidé à son élaboration.

On retrouve ici le phénomène bien connu de fétichisme du chiffre, c'est-à-dire ce qui se produit lorsque les chiffres (ou les lettres !) circulent et s'imposent dans l'espace public, indépendamment de leur procédé de fabrication. Dans ce cas, le chiffre échappe à son créateur pour vivre une existence autonome dans l'espace social. Les conditions sont alors réunies pour qu'il soit utilisé, non comme un élément de connaissance mais plutôt comme un argument d'autorité. En ce sens, la généralisation d'évaluations chiffrées représente, non un progrès mais plutôt un danger pour la démocratie : d'une part parce que ces évaluations occultent tout ce qui n'est pas facilement mesurable ; d'autre part parce qu'elles risquent d'être utilisées pour clore des débats qu'il importe, en démocratie, de maintenir ouverts. Un peuple d'indicateurs s'est progressivement imposé dans notre monde social : s'il se présente sous la forme d'avatars diversifiés et individuellement vulnérables, la machinerie reste soudée par un même esprit de classement et de comparaison systématique. Ce n'est pas ici l'opération de quantification qui est en cause mais bien l'autonomisation de ses produits, légitimée par des idéaux de justice et de transparence démocratique.

Annexe 6

Réflexions croisées sur la mesure et l'incertitude

François Vatin
Alain Caillé
Olivier Favereau

Revue française de gestion, n°203, 2010/4

-Extraits-

Nota : toutes les notes de bas de page ainsi que la bibliographie du texte ont été supprimés.

Certains passages ont été soulignés par nous pour attirer l'attention du lecteur et faciliter une lecture rapide.

Contribution de François Vatin

L'incertitude du monde social et la raison métrologique : une approche pragmatiste

Comme on le sait depuis Condorcet, l'agrégation des choix individuels pour construire une fonction de choix collective débouche sur des apories insolubles.

La mesure est réductrice. C'est là sa fonction même. On ne produit de l'information qu'en en détruisant. Il s'agit de réduire l'infini chatoisement du réel à un cadre métrologique qui en rend compte de façon simplifiée.

Cela commence avant même l'attribution de valeurs chiffrées par les catégorisations : dans quelle classe range-t-on tel élément ?

La mesure du sociologue, comme celle du gestionnaire n'échappe pas à cette incomplétude principielle : toute mesure est partielle, partiale, inconséquente, provisoire, "fausse" en toute généralité. Faut-il pour autant y renoncer? C'est là qu'il faut introduire un point de vue pragmatiste, qui permet de mettre en évidence le fait que la mesure a moins pour vocation à dire le monde qu'à agir sur lui. La mesure est le produit des "ingénieurs", pas des savants. On peut résumer, au risque de la caricature, la pensée des ingénieurs par ces deux adages : "mieux vaut une mauvaise mesure que pas de mesure du tout"; "on mesure ce que l'on peut mesurer". Pour s'orienter dans le monde, il faut des points d'appui ; il faut bien les prendre où on les trouve. Confrontée au réel, la mesure montrera elle-même ses limites, appellera à son renversement paradigmatique, quand d'autres mesures plus fines seront disponibles. En attendant elle nous sert, car nous resterions sans elle aussi indéterminés que l'âne de Buridan. En poussant à ses limites cette parabole, une mesure totalement arbitraire resterait préférable à l'absence totale de mesure. Autrement dit, la mesure, loin de nous faire perdre de vue le caractère fondamentalement incertain du monde, nous le rappelle. C'est parce que le monde est incertain que la mesure est nécessaire

Platon défend une conception utilitariste de la Cité et de la justice, fondée sur un idéal de la mensuration, et Aristote une conception anti-utilitariste qui se refuse à la quantification de nos actes et du rapport social.

Selon la philosophie de chacun on pourra lire dans cette évolution un changement bien venu, permettant enfin d'en finir avec un ordre mandarin archaïque, fait de petits arrangements entre pairs, de miniprébendes et d'assez grande inefficacité globale. Ou, au contraire, y voir le coup de boutoir ultime porté contre l'*ethos* du savoir plus ou moins désintéressé et la mort programmée de toute pensée réflexive autonome.

Pour la première perception la systématisation de la quantification de la valeur scientifique donne un coup de pied salutaire dans la fourmilière, elle révèle des choses bien cachées, les rentes de situation, la complicité des fainéantises ou des médiocrités.

Aussi incertaine que soit la mesure elle vaut à coup sûr mieux que pas de mesure. Pour la seconde perception, l'imposition de normes comptables absurdes aboutit à tuer la poule aux oeufs d'or en démotivant les vrais savants pour ne favoriser que les arrivistes et les truqueurs, ceux qui savent se faire bien voir des métreurs et autres géomètres.

Seule une certaine incertitude sur la valeur relative des hommes, de leurs actes et de leurs productions autorise le maintien d'une société ouverte et démocratique.

En tout état de cause, précisément parce que la valeur relative des divers types de savoir est indéterminable, parce que leurs composantes sont nécessairement plurielles, l'idée même qu'il serait possible d'arriver, fût-ce approximativement, à une bonne mesure unique, se révèle totalement intenable.

Si la grève des Universités a revêtu à l'hiver 2009 l'intensité que l'on sait, c'est parce que les réformes imposées par le gouvernement semblaient viser, *via* l'obligation de se soumettre à des évaluations quantitatives répétées, et sous prétexte d'autonomie des Universités, à déposséder définitivement la communauté savante de la part d'autonomie qu'elle conservait. Ce n'est donc pas l'existence d'une norme métrique qui fait en elle-même problème. Les professeurs qui ont passé toute leur jeunesse à rechercher de bonnes notes et qui continuent à noter à tour de bras seraient mal placés pour contester toute quantification de la valeur des productions scientifiques. Ce qui est problématique n'est donc pas l'introduction de la mesure en elle-même mais deux choses assez étroitement liées : le fantasme de la bonne mesure unique (du *One Best Way*) et la liquidation de l'autonomie relative des corps professionnels qu'il favorise.

Du coup, et pour rejoindre peut-être François Vatin, je serais tenté de conclure en disant : non à *la* mesure de la valeur du savoir, mais oui à une multiplicité des mesures, qui laisse la question de la valeur ultime des uns et des autres dans l'incertitude et qui permette à la communauté universitaire – ou à d'autres communautés selon la même logique – de reconquérir son autonomie perdue.

Contribution d'Olivier Favereau
L'incertitude sans mesure

La théorie économique dominante, justement parce qu'elle valorise à l'extrême le calculable, qu'il s'agisse de la rationalité individuelle, ou de la coordination marchande, réduit systématiquement l'incertitude de l'avenir à un calcul de probabilités.

Le propre de l'incertitude, c'est "on ne sait pas".

Voici le sens ultime de l'incertitude, dès lors qu'elle est radicale : l'avenir n'est plus simplement à prévoir, il est aussi à provoquer ; il n'est plus simplement à découvrir, il est aussi à inventer. Si l'action – la gestion – doit être portée par une vision du long terme, c'est que le fait d'agir dans cette direction contribuera à la rendre effective.

L'incertitude radicale sur la nature du produit "recherche". Elle s'alimente à deux sources, qui ne sont pas indépendantes : qui donc est demandeur de la recherche ordinaire, celle qui va se traduire par des publications qui disent d'elles-mêmes qu'elles sont "scientifiques" et comment juger de la quantité et de la qualité de la recherche produite par *l'ensemble* d'une Université ? Difficile de porter un jugement *d'ensemble*, qui requerrait une compétence exceptionnelle. On va donc descendre à un niveau plus micro, celui des chercheurs eux-mêmes, en écartant au passage les interactions entre chercheurs – si on ne les écartait pas, il faudrait porter un jugement *d'ensemble* sur la recherche produite au niveau (probablement) du laboratoire – et on serait renvoyé à l'impasse précédente, sous une forme à peine moins aiguë. Donc on finit par se résigner à cette forme... d'individualisme méthodologique extrême dans l'évaluation de la recherche : le comptage des publications, dans les revues dites "scientifiques".

Le problème était de faire sortir le produit de la recherche de son indétermination foncière, notamment en essayant de déterminer quel acteur économique peut bien se porter demandeur d'un tel produit. La réponse s'impose d'elle-même dans sa crudité : ce sont les revues académiques – donc des entreprises – qui incarnent la demande "finale" des produits de la recherche. Au même titre que les consommateurs pour les autres produits, les revues décident, par leur acceptation ou leur refus des projets d'article, à la fois de la recherche admissible, mais aussi de la valeur de toute recherche.

Est-ce tout à fait la demande "finale" ? Il semble que non : les entreprises que sont les revues académiques doivent vendre les numéros qu'elles fabriquent. À qui ? Essentiellement aux Universités ! Les Universités payent les entreprises-revues qui (dé) valorisent leur contribution à la recherche. Le système est refermé sur lui-même.

Il est temps d'ouvrir au principe de réalité un système qui s'est fermé sur lui-même : les publications scientifiques ne sont pas le produit final de la recherche, mais un produit intermédiaire, certes essentiel, mais encore en attente des opérations, individuelles et collectives, qui le transformeront peut-être en un surcroît d'intelligence pour tous.

CONCLUSION

Le traitement de l'incertitude par la mesure peut alors aboutir à une forme de captation de pouvoir cognitif susceptible de générer des rentes économiques.

Parce qu'ils sont plongés dans l'histoire, le savoir, comme toute production sociale, ne sauraient être réduits à une métrologie, si raffinée soit-elle. La métrologisation généralisée à dominante économique qui tend aujourd'hui à s'imposer n'est donc pas sans risques pour la cohésion sociale présente, mais aussi pour la capacité des sociétés à se penser dans un futur.

Le présent débat sur l'Université souffre tout à la fois de l'insuffisance gravissime de mesures et de réflexion métrologique pour penser l'efficacité de nos systèmes d'enseignement et de recherche et d'une illusion tout aussi grave sur l'idée que la question pourrait être réduite dans une métrologie économique universelle, qui est au final incapable de penser le savoir dans ce qu'il a de plus spécifique : la production de "nouveau".

Annexe 7

Revue qui compte, revue qu'on compte : produire des classements en économie et gestion

David Pontille
Didier Torny

Revue de la régulation
Numéro n°8 (2e semestre 2010)

-Extraits-

Nota : toutes les notes de bas de page ainsi que la bibliographie du texte ont été supprimés. Certains passages ont été soulignés par nous pour attirer l'attention du lecteur et faciliter une lecture rapide.

Peut-on (...) trouver un ordonnancement consensuel pour ces groupes de revues, à défaut d'un classement individualisé ? Moore (1972) a répondu positivement à cette question en produisant un classement à cinq niveaux, fondé sur le prestige des affiliations institutionnelles des auteurs d'articles. Il considérait que ce classement, malgré quelques imperfections, devrait remplacer les points de vue partiels et partiels, afin de permettre une meilleure allocation des ressources sur le marché des revues. En revanche, Hawkins et al. (1973) s'y sont refusés, estimant *in fine* qu'il n'est pas possible de produire une évaluation objective des revues.

Les articles portant sur les classements de revues en économie ou en gestion constituent ainsi, à partir des années 2000, un sous-domaine académique, qui fourmille de productions de metaclassements.

En retour, ces classements, et particulièrement ceux des "meilleures revues", sont utilisés pour mesurer la qualité des *departments* ou des *business schools* via le nombre de publications de leurs membres.

Étudier les classements de revues en économie et gestion présente donc un premier intérêt : celui d'observer cette activité dans des disciplines où elle est récurrente. L'existence d'une littérature beaucoup plus vaste que dans d'autres disciplines amène à poser la question de leur singularité : les économistes et gestionnaires déploient-ils des instruments spécifiques, qu'ils soient de l'ordre de la théorie, de la méthodologie ou du calcul ? Ou s'agit-il simplement d'une croyance partagée en la possibilité de valoriser et d'évaluer, de compter et de classer ? Ces listes ont été utilisées pour évaluer la production scientifique des Universités à partir de 1995.

On sait désormais que les classements peuvent modifier l'ensemble des pratiques de recherche et de publication. Ce phénomène ne surprend guère les économistes et gestionnaires, quel que soit leur modèle théorique : maximisation de l'utilité, force des conventions partagées, puissance des technologies invisibles... et il n'y a aucune raison pour que la production des économistes et gestionnaires y échappe.

Étudier les classements de revues dans ces disciplines présente donc un deuxième intérêt : celui d'analyser la prise en compte par les acteurs de la dimension performative de ces

instruments, qu'elle soit considérée comme une simple explicitation des valeurs déjà partagées ou comme une assignation nouvelle.

Aussi, étudier les classements de revues dans ces disciplines présente un troisième intérêt : les acteurs concernés mobilisent des savoirs, sinon plus spécialisés, au moins segmentés d'une manière beaucoup plus marquée que dans la plupart des autres SHS. Nous l'avons vu, dans le cas de Hawkins et al. (1973), le type de département universitaire modifiait radicalement la position de certaines revues dans le classement.

Il est inutile ici de rappeler en détail la domination d'une économie *mainstream* sur d'autres courants théoriques, mais il faut ajouter l'importance des asymétries entre les productions nord-américaine et européenne, y compris en gestion. Plus que pour d'autres SHS, les négociations autour des classements de revues mettent en jeu la définition et les limites des groupes professionnels.

Pour ces trois raisons, l'économie et la gestion sont des milieux où l'explicitation des procédures de fabrication, d'anticipation des usages et de clarification des enjeux est maximale. S'y intéresser permet de rendre compte de l'ensemble des positions possibles, des arguments qui les soutiennent et des dispositifs qui font l'accord ou le désaccord.

Dans un contexte où les classements de revues se sont multipliés, la section 37 du CNRS, composée d'économistes et de gestionnaires, s'est lancée dans la production de son propre classement. Poursuivant un travail non abouti durant la mandature 1997-2000, la section a formé en 2001, sous l'impulsion du président de la nouvelle mandature, une "commission revues" dont l'objectif était sans équivoque : produire un classement à usage interne, pour l'évaluation des candidats aux concours et des unités de recherche, prenant donc la forme d'un "outil d'aide à la décision plus que celle d'un outil absolu de classement qui donne des résultats aberrants et ôte tout besoin d'une instance d'évaluation et de jugement scientifique.

Pour autant, le classement provisoire de 487 revues diffusé fin avril 2003 n'a pas fait consensus. La contestation a dépassé le cadre confiné d'une discussion disciplinaire.

Trois limites du classement des revues en gestion ont été épinglées. Il était d'abord jugé discriminatoire sur le plan linguistique puisqu'il valorisait principalement "un petit nombre de revues de langue anglaise" et que les membres de la section avaient eux-mêmes négligé le français, "comme d'ailleurs toutes les autres langues possibles de publication", en tant qu'instrument de communication. La section avait anticipé cette critique en traitant en amont ce qu'elle appelait "l'exception française" :

Mais (pour M. Berry et certains collègues), la discrimination linguistique ne pouvait être isolée du contenu même des recherches. Ils pointaient ainsi le risque d'une standardisation des travaux par l'alignement sur des critères internationaux, en fait nord-américains, dont les vecteurs étaient les revues bien classées par la section.

Loin de penser qu'il devait nécessairement être élitiste, les promoteurs du classement le concevaient avant tout comme un appui conventionnel, à la fois lieu d'inscription de normes professionnelles partagées et instrument de justice sociale donnant à chacun les moyens d'argumenter ses positions.

Deux conceptions s'affrontaient donc autour du même dispositif : la première considérait le classement comme une simple objectivation des états précédents du monde, dont l'usage est par nature circonscrit et dont les effets éventuels sont recherchés (diminution des asymétries d'information, égalité de traitement des personnes, et amélioration du niveau de publication des chercheurs), tandis que la seconde traitait le classement comme une "technologie

invisible" qui crée de nouvelles hiérarchies et standardise les pratiques et les jugements (marginalisation de certaines revues, diminution de la diversité des méthodes et des objets).

En dépit d'une argumentation très centrée sur la langue d'expression des revues et la défense de leurs "terroirs de recherche" on n'a pas assisté à un vaste mouvement collectif des revues françaises. C'est plutôt sous la forme de demandes individuelles que quelques revues se sont adressées à la section en pratiquant un plaidoyer *pro domo*. Elles ont notamment fait valoir la rigueur de leurs procédures d'évaluation et ont exhibé des éléments quantitatifs pour figurer à une meilleure place.

Les instigateurs ont terminé leur mandat sur une réussite : ils ont bel et bien stabilisé un classement de revues. À l'image de nombreuses autres innovations technologiques, son caractère un peu bricolé a suscité des critiques, auxquelles il a néanmoins résisté.

Le classement de revues a été intégré par les grandes écoles de commerce en tant qu'outil d'évaluation de la recherche.

Ce qui a été conçu comme une aide au jugement par les pairs est donc devenu un outil de calcul direct pour une institution, aussi concret que le nombre d'étudiants par enseignant permanent.

Les nouvelles catégories d'évaluation ont rapidement été reprises par les évalués, déclenchant ce que certains ont appelé "la course aux étoiles". Estimant qu'ils n'étaient dorénavant plus jugés sur la lecture directe de leurs travaux, ni même sur le proxy de qualité que constitue le nom d'une revue, certains laboratoires ont organisé leur rapport autour du nombre d'étoiles.

Deux ans après sa publication, le classement, largement discuté, débattu et repris par d'autres institutions, faisait désormais partie du paysage.

Annexe 8

External review report of the French Evaluation Agency for Research and Higher Education (AERES) with regard to the Standards and Guidelines for Quality Assurance in the European Higher Education Area, May 2010.

Critiques formulées à l'AERES par l'ENQA (European Association for Quality Assurance in Higher Education)

- Extraits-

p. 37

“ 5.2 Summary of the main points to improve

- The AERES must continue its efforts to develop an evaluation culture within institutions by paying greater attention to the quality of the self-evaluation provided by institutions and to the participation of professors, students and staff in producing it.
- Along the same lines, the AERES should improve its evaluation guide by incorporating criteria focusing specifically on the quality assurance strategy of institutions or by making existing criteria more precise, and ensure that the means provided for in this strategy are put into practice.
- The AERES' procedures for evaluating bachelor's degrees and master's degrees should be revised to bring them more into line with the ESG requirements.
- The AERES' strategic plan has its positive points, but could be improved by the agency specifying the means it intends to implement to achieve its objectives, the persons responsible for this and the timeframes.
- In its strategic plan, the AERES wants to add a greater international dimension to its activities. The panel encourages it to do so, amongst other things by calling more systematically on international experts to carry out its evaluations.
- With the current evaluation cycle, there is a risk that evaluations become routine and ineffective. The panel suggests that the AERES look into the possibility of extending the current cycle.
- Alongside its Council, the members of which are stipulated by Law, it could be in the AERES' interests to set up an advisory committee with members from various sectors – particularly students and international specialists.
- The AERES has excellent Quality Standards. It would be worth linking these explicitly in with its evaluation guide.
- With this in mind, the panel suggests that the AERES make the criteria in its evaluation guide more precise and ensure that they concern both the effectiveness and existence of quality assurance procedures and policies.
- The AERES' procedures are promising. They could be improved from several aspects:
 - ♦ The AERES could send the preliminary version of its evaluation report to institutions to obtain their comments before writing the final report.
 - ♦ This stage should not, for all that, change the procedure in place of integrating the institution's reactions to the final report.

- The AERES could also send the final version of the report to the experts before it is put on its website.
 - The question of scores and their publication raised several comments during the site visit. The AERES should perhaps discuss this issue with the main stakeholders – particularly in the research sector. It would also do well to consider the possibility of revising the score when clear improvements have been made.
 - The AERES would gain from annually updating its pool of student experts.
- The AERES should set up follow-up procedures to enable it to assess the measures taken following its evaluations quickly”

p. 17

“The Quality Standards address the same criterion by specifying the key points to be examined: *“Existence of a policy guaranteeing the quality of programmes and degrees and divided up into operational procedures; definition and implementation of a quality continuous improvement strategy; publication of the policy, strategy and procedures; actual role of students and stakeholders in the quality management system.”* Eight specific criteria follow, covering all of this standard’s points, for example the institution *“has implemented quality assurance and assesses and analyses its effectiveness (achievement of objectives).”* This type of criterion goes further than those found in the guide. It would be in the AERES’ interests to draw on this when producing its next guides. To sum up, the AERES complies substantially with this standard, **but could improve its services through more precise and binding criteria”**.

p. 18

“The evaluation guide goes suitably into this question in objective 3.5 on the teaching quality system. The quality of student evaluation must be examined from the point of view of examinations designed as a pedagogical procedure, updated evaluation criteria and procedures based on pedagogical knowledge, the existence of clear rules and criteria, etc. **However, the institution’s role in checking that procedures are applied and criteria are followed is not mentioned”**.

p. 18

“The role of students in the evaluation is taken into account. The management of teachers’ and professors’ jobs, recruitment and use of resources are brought up in objective 7.3. **However, nothing is said about improvement measures”**.

p. 20

“To sum up, the AERES’ evaluation criteria take on board all of the standards of Part 1 of the ESG. In this regard, the AERES is in line with standard 2.1 of the ESG. **However, its criteria are described in much more detail in the Quality Standards it adopted in December 2008 than in its evaluation guide.** At present, the criteria of the evaluation guide are expressed in question form, which might raise interpretation difficulties. The experts’ work would be facilitated by a text that is more precise and more in line with international practices. The expert panel asks the AERES to improve its evaluation guides by looking to the ones it developed itself in its Quality Standards. Moreover, in several cases, whilst the criteria seek to ensure the existence of policies, regulations or procedures, they do not actually state

any measures to check that these are effective. This is the case, for example, in objective 7.7 on quality assurance, objective 3.3 on student support initiatives and objective 3.5 on the teaching quality system. This does not mean that the reviewers cannot check the effectiveness of these policies, regulations and procedures – but the guide does not prompt them explicitly to do so. Here again, the Quality Standards go further and could be used as a starting point for making improvements”.

p. 22

“Some of the people interviewed during the visit nevertheless pointed out that there should preferably have been greater consultation on the evaluation criteria. However, the AERES’ efforts in disseminating the evaluation concepts and method should be commended. For example, before evaluating each group it organizes meetings with the institution heads, their executive teams and the directors of research units”.

p. 23

“The Panel considers that the AERES complies adequately with standard 2.3 of the ESG in terms of evaluations of institutions and research units. Improvements could be made regarding the evaluation of training programmes to make the decision procedure more explicit. The expert panel will come back to the subject of evaluating training programmes in the context of analysing conformity to ESG 3.7”.

p. 24

“Overall, the AERES is in line with this standard when taking purely evaluations of institutions, research units and doctoral schools into consideration. But the Panel has reservations over the evaluation of the quality of bachelor’s degrees and master’s degrees”.

p. 25

“The panel commends the AERES’ efforts in making its evaluation reports accessible. The panel was somewhat intrigued by the attribution of scores for training programmes and research units. It seems that this has been practised in France for a number of years now and is relatively well accepted.

Nevertheless, some of the people interviewed during the site visit expressed reservations over grading, and particularly the use that is made, by external stakeholders, of the scores attributed. It is not the panel’s place to contribute to the discussion on this matter. That said, the panel points out that the publication of a score – especially when it is unfavourable, risks oversimplifying the evaluation findings of the research unit or training programme. Above all, maintaining this score for four years risks unfairly qualifying the unit or programme once it has made the recommended improvements. The expert panel suggests that the AERES re-examine this publication practice”

p. 26

“The AERES fully conforms to this standard. However, for the reasons mentioned above, the expert panel strongly recommends that the AERES and its partners look into the possibility of extending the frequency with which evaluations are carried out”.

p. 29

“The AERES has also set itself objectives that are presented in different ways in its various documents. These may be summarised as follows: implement the European commitments made by France through the Bologna Process; assist the evaluated institutions so as to improve their governance, research and range of programmes; provide the State with an impartial tool for making its strategic decisions and provide students and other stakeholders with useful information.

Similar objectives can be found in most national agencies and, generally, they give precedence to the assistance they can give to institutions. In its evaluation guide, the AERES also indicates that it “*puts the institution at the heart of the evaluation process, the purpose of which is to help the latter to undertake a continuous improvement policy.*” Nevertheless, because the AERES’ mission is tied in with the contract establishment process, the objective of providing the State with an impartial tool stands particularly out – such that, for some of the people interviewed during the site visit, this objective is considered to be the most important. Were this opinion to become the norm, there is a danger that the relationship of assistance that the AERES wishes to forge with universities would lose much of its effectiveness. Whilst fully aware that the AERES must comply with the terms of its founding Decree, the panel believes that the AERES should take care as regards this possibility”.

p. 31

“These legal provisions are sufficient to guarantee the AERES’ independence. However, it should be noted that the AERES’ ability to define its work programme is limited by the requirement laid down in its founding Decree to adopt a multiannual evaluation programme that is “*compatible with the procedural deadlines for drawing up contracts between institutions and the State*”. As such, the AERES is not completely free to develop its work programme, as has already been mentioned. Incidentally, it is this requirement that has led the AERES to organize its evaluation

programme according to a four-year cycle, even if, by its own admission, this cycle places a workload on it that is difficult to keep up in the medium term. This is a constraint it must abide by, but it cannot be said that it calls the AERES’ independence into question. The AERES has chosen its procedures, criteria and regulations itself. Neither does this constraint affect its ability to make its own choices of experts”.

Annexe 9

Le danger des mauvais usages des indicateurs

VRS 374, Juillet/septembre 2008

La critique qui semble la plus complète sur cette question de la bibliométrie semble venir d'Yves Gingas, statisticien québécois, qui a écrit un article absolument définitif intitulé "Du mauvais usage de faux indicateurs", paru en 2008 dans "La revue d'histoire moderne et contemporaine" et repris dans la VRS de septembre 2008.

Il montre entre autres le détournement complet de la bibliométrie, initialement conçue par une petite communauté de bibliothécaires, sociologues, statisticiens qui étudiaient les transformations de la recherche à travers les propriétés des publications et les références qu'elles contiennent. Il est évidemment passionnant et très utile de savoir sur quoi les gens travaillent, quelles sont les thématiques qui émergent, celles qui stagnent ou régressent, les contestations, mais aussi qui travaille et publie avec qui etc. Dans cette approche, il faut se saisir des questions, et c'est essentiel, à un niveau agrégé (grands laboratoires, Universités, pays) et surtout pas au niveau individuel. En aucun cas utiliser ces indicateurs pour évaluer les individus, en particulier au recrutement et s'agissant de jeunes. On est loin des usages qu'on fait actuellement de la bibliométrie.

Annexe 10

Réflexion sur l'évaluation et les listes de revues. L'exemple de l'économie et de la gestion

Florence Audier

La vie des idées, 15/09/2009

Résumé : L'évaluation, et la place qu'y occupe la bibliométrie, sont au cœur des débats et même parfois des affrontements actuels. Les enjeux sont majeurs. Va-t-on déléguer aux directions des revues, à leurs *boards* et leurs *referees*, ainsi qu'à leurs clients et lecteurs le pouvoir de trier parmi les recherches – en déterminant celles qui sont dignes d'intérêt et celles qui ne le sont pas – et parmi les chercheurs – en distinguant les "excellents" des autres. Car il ne faut pas s'y tromper : avec la bibliométrie telle qu'on cherche à nous l'imposer, ce sont les supports de publication – en l'occurrence les revues, au détriment d'ailleurs des autres modes de diffusion du savoir, tels les ouvrages – auxquels on transfère *ipso facto* le pouvoir exorbitant d'évaluer, l'AERES, l'ANR, et bientôt les dotations en crédits des Universités considérant la bibliométrie comme l'indicateur suprême de l'activité et de la qualité de la recherche. D'où l'intérêt d'examiner les choses de façon concrète.

(...) Après avoir réfléchi aux relations aléatoires entre la contribution aux revues et la qualité des recherches, et montré que la politique qui cherche à s'imposer est éloignée de celle qui est prétendument appliquée aux États-Unis, l'article présente les choix effectués par la section "Économie/gestion" du Comité national en matière de "liste de publication", et les considérants qui les accompagnent. Enfin, on détaille les résultats d'une étude centrée sur *la fréquence d'accès des auteurs français aux revues classées par la communauté des économistes français comme les meilleures et érigées en critères d'excellence*. Cette étude, précise, qui porte sur 1636 articles et 3191 signataires ou cosignataires, s'intéresse à la fois aux types de contributeurs et aux domaines retenus par les 57 titres de revues considérées par la communauté comme de "top niveau", apporte la démonstration que les revues classées ici comme étant de rang 1 (1* pour les revues dites "généralistes") ou A, toutes anglo-saxonnes, ne publient, en réalité, pratiquement que des universitaires appartenant à leurs cercles les plus proches, et que les contributeurs français en sont quasi totalement absents (sauf dans deux sous-disciplines spécifiques). Les français ont donc désigné comme cible d'excellence des revues... auxquelles ils n'accèdent pas, sauf réseaux particuliers construits avec des collègues américains ou assimilés ayant déjà pied dans ces revues en tant que membres d'un *board*, comme *referees*, ou comme *contributeurs* habituels.

-Extraits -

Si l'on réfléchit à présent aux incidences de ces choix sur la recherche elle-même et son devenir, il faut tout d'abord noter quelques points "positifs" : l'examen des sommaires et des *abstracts* permet en effet d'identifier relativement rapidement quels sont les sujets traités - à un moment donné - par les chercheurs (du monde entier ? on reviendra plus loin sur le sujet de l'accès à ces titres et de leur diffusion), et de dialoguer éventuellement avec eux, de les contacter (et plus si affinités...), bref de se repérer dans cet univers complexe qu'est notre "communauté scientifique". Cela permet en même temps de savoir quels sont les sujets "à la mode" - *i.e.* ceux qui ont été retenus par les revues patentées- ce qui peut être d'un grand secours pour postuler auprès des diverses "agences de moyens". Enfin, dans le même ordre

d'idées, cela permet de se repérer dans le maquis des supports, pour identifier sans risques "ce qu'on fait de mieux" et tenter de s'y conformer pour y accéder ou y faire accéder nos doctorants.

(...) Pourtant, ce système recèle de graves défauts qui sont l'exact envers de ces avantages, mais qui vont bien au-delà. On voudrait ici insister sur un aspect insuffisamment évoqué, et qui ne se retrouve pas – ou moins – dans les sciences "exactes" : *le conformisme que ce type d'évaluation génère concernant les thématiques, les paradigmes, l'origine et le type de données mobilisées, les plans d'expositions et d'argumentation, et donc les façons de réfléchir, d'écrire et de convaincre, défauts qui annihilent largement les avantages potentiels de la forme "article"*. Les exemples pullulent : pour avoir une chance d'être publié, il faut tout d'abord lire attentivement les conseils aux auteurs et s'imprégner des articles des autres, pour identifier puis entrer dans le moule. Si l'on est européen, et singulièrement issu d'un pays du Sud, pour avoir une chance d'être publié, mieux vaut trouver un co-auteur anglo-saxon, ayant déjà publié dans une des revues répertoriées dans la "top list", ou, mieux encore, appartenant à un *board* (ce sont d'ailleurs souvent les mêmes). Ce sera de préférence un universitaire qu'on a connu lors de son post-doc, ou bien que son "patron" a fait venir comme professeur invité (ou réciproquement) ou qu'on a croisé dans un colloque où votre équipe vous a fait la faveur de vous propulser (et de vous financer). Pour avoir une chance d'être publié, mieux vaut être *supporté* en affichant que le papier soumis a déjà été de multiples fois présenté lors des cérémonies que sont les grands colloques ou congrès rituels aux Etats-Unis, qu'il a été lu et relu - donc que l'auteur a été adoubé - par des personnalités prestigieuses (dont on cite abondamment les noms et qualités sous forme de remerciements, généralement en note de première page pour être sûr que ce sera lu), que la recherche ici publiée a bénéficié de *grants* dont les donateurs sont aussi remerciés, et a été formidablement amélioré par les pertinentes remarques des rapporteurs, etc... bref, qu'en jugeant – et éventuellement en refusant – un papier, c'est toute cette communauté sus-nommée qui est jugée. Non pas seulement parce qu'elle est supposée avoir contribué à "améliorer" le texte, comme il est souligné en exergue, mais parce qu'elle s'est en quelque sorte engagée en le patronnant. (...) Lorsque la trouée est faite, on peut amortir ses efforts, déjà soi-même, en publiant plusieurs articles de suite, et en facilitant l'accès à d'autres (à charge de réciprocité ?), d'où parfois des publications "en grappe" de membres des mêmes équipes dans les mêmes revues, sans que des raisons, par exemple thématiques, ne semblent déterminantes.

Tableau récapitulatif de l'accès des "français" aux revues classées 1* pour les revues généralistes et 1 pour les revues classées par domaines

Domaine	Nombre d'articles	Nombre d'auteurs	Nombre d'auteurs français[°]	Nombre d'articles signés par des français[°]
Revues généralistes	<i>316</i>	<i>630</i>	<i>17</i>	<i>11</i>
Développement et transition	<i>153</i>	<i>290</i>	<i>7</i>	<i>4</i>
Droit et économie	<i>91</i>	<i>166</i>	<i>0</i>	<i>0</i>
Finance et assurance	<i>167</i>	<i>345</i>	<i>1</i>	<i>1</i>
Gestion des R H	<i>56</i>	<i>118</i>	<i>0</i>	<i>0</i>
HPE, Histoire économique (hors la revue des Annales)	<i>99</i>	<i>124</i>	<i>15</i>	<i>12</i>
Macro économie, économie intern.	<i>108</i>	<i>208</i>	<i>4</i>	<i>4</i>
Organisation industrielle	<i>97</i>	<i>185</i>	<i>4</i>	<i>3</i>
Économie publique	<i>115</i>	<i>216</i>	<i>16</i>	<i>10</i>
Économie de la santé	<i>112</i>	<i>329</i>	<i>3</i>	<i>2</i>
Économie spatiale, transports	<i>73</i>	<i>139</i>	<i>3</i>	<i>2</i>
Théorie éco, économétrie, jeux	<i>179</i>	<i>308</i>	<i>21</i>	<i>11</i>
Économie du travail	<i>70</i>	<i>133</i>	<i>2</i>	<i>2</i>
<i>Ensemble des revues classées 1 ou 1* (pour GEN)</i>	<i>1636</i>	<i>3191</i>	<i>93</i>	<i>62</i>

[°] sont considérés ici comme des auteurs français les auteurs de toutes nationalités dont l'institution accolée à la signature est française. Le dépouillement couvre environ deux années

Annexe 11

Une étude comparative internationale des bases de données des revues scientifiques en sciences humaines et sociales (SHS)

Michèle Dassa

Kristine Kosmopoulos

JournalBase

Il s'agit d'une étude comparative des bases de données suivantes en SHS (toutes disciplines) :

- WOS-AHCI et WOS SSCI : **3936 titres de revues** (dont 1885 AHCI et 2642 SSCI)
- Scopus SHS : **4503 titres**
- Listes ERIH (European Reference Index for the Humanities) : **5131 titres**
- Listes AERES (SHS) : **6245** (listes droit et sciences de la communication non intégrées) ;

Au total, **la base JournalBase** (qui résulte de la fusion) 2010 contient **10756 titres (9551 en 2009)**. **L'intersection entre ces quatre bases n'est que de 9,6%, c'est-à-dire que 9,6 %** des titres figurent dans les 4 bases, d'où un problème majeur de représentativité des travaux publiés dans les revues consignées dans les unes et les autres.

Le nombre de revues françaises par bases est le suivant : AERES 427, ERIH (Europe) 318, SCOPUS 154, WOS-ssci 29.

En "économie", Journalbase contient 997 revues, 55 sont françaises (5,5%), 3 des 39 revues communes aux différentes bases relèvent de l'économie

Annexe 12

Rapport de l'Académie des Sciences

Du bon usage de la bibliométrie pour l'évaluation individuelle des chercheurs

Jean-François Bach, Secrétaire perpétuel de l'Académie et coordonnateur du rapport

Denis Jérôme, Membre de l'Académie et coordonnateur du rapport

Contact : Brigitte d'Artemare (brigitte.d-artemare@academie-sciences.fr)

L'Académie des sciences a remis le 17 janvier 2011 à la Ministre de l'Enseignement supérieur et de la Recherche un rapport intitulé *Du bon usage de la bibliométrie pour l'évaluation individuelle des chercheurs*.

Synthèse-résumé-recommandations :

L'évaluation bibliométrique est fondée sur l'utilisation des citations des articles scientifiques pour en estimer l'impact et souvent la qualité. Elle est rendue possible par l'existence de bases de données informatiques qui permettent, pour chaque chercheur, de calculer des indicateurs à partir des citations de leurs articles.

La bibliométrie est largement utilisée dans de nombreuses disciplines pour l'évaluation des individus ou des institutions. Néanmoins, cette utilisation fait débat avec des avis très tranchés positifs ou négatifs, en grande partie en fonction des disciplines, mais pas seulement. C'est dans ce contexte que l'Académie des sciences a souhaité se pencher sur ce problème, avec toute la mesure possible, en tenant le plus grand compte des différences culturelles et des pratiques de chacune des disciplines sur le sujet.

L'outil bibliométrique doit son succès à son apparente simplicité d'utilisation, contrastant avec la lourdeur et certaines limites de l'évaluation qualitative par les pairs. Il associe des nombres à chaque chercheur ce qui est attrayant car cela permet une première évaluation rapide fondée sur des données factuelles et objectives. De fait, la bibliométrie peut être utile pour une instance d'évaluation afin de réaliser un premier tri et se faire une opinion rapide. Néanmoins, l'outil bibliométrique expose à des biais sérieux quand il n'est pas utilisé de façon adéquate. En premier lieu, il ne peut en rien remplacer l'évaluation qualitative par les pairs pour lesquels il pourrait cependant représenter un outil intéressant. **Seule l'évaluation qualitative peut apprécier l'originalité et l'esprit d'innovation d'un chercheur. Cette réserve majeure implique que la bibliométrie ne doit être utilisée que par les pairs, et surtout pas par des personnes n'en ayant pas l'expertise, ou n'ayant pas pris le temps nécessaire pour examiner les dossiers scientifiques dans leur ensemble.** En outre, sur un plan plus technique, l'outil bibliométrique n'a de valeur que si certaines conditions d'utilisation sont respectées, ce qui malheureusement est loin d'être la règle. Il faut d'abord s'assurer que les calculs des indices bibliométriques ont été correctement réalisés. De plus, il faut savoir que dans les disciplines où l'ordre des auteurs n'est pas alphabétique mais tient compte de la contribution de chacun d'entre eux, la place de chaque auteur dans un article doit être prise en considération, ce qui n'est habituellement pas fait. Surtout, et ce point est crucial, il n'est pas possible de comparer les indices bibliométriques de chercheurs appartenant à des disciplines ou même des sous-disciplines différentes. Il s'avère en effet que les valeurs des indices bibliométriques varient beaucoup selon les disciplines, au point qu'il doit être fortement recommandé de se référer, pour chacune d'entre elles, à la distribution des indices autour de la moyenne, en tenant compte de l'âge des chercheurs.

Pour toutes ces raisons, sans vouloir être trop critique sur l'utilisation de la bibliométrie et en sachant qu'elle est et sera en tout état de cause largement utilisée, de façon transparente ou cachée, **l'Académie insiste sur l'importance du respect des règles qui viennent d'être énoncées, pour les disciplines qui s'y prêtent ce qui exclut aujourd'hui les mathématiques et les sciences humaines et sociales.** Pour approfondir cette discussion, l'Académie propose la création d'un comité de pilotage placé auprès de l'Observatoire des Sciences et des Techniques (OST), qui serait chargé de valider par des tests rétrospectifs et d'autres études, les indicateurs actuellement utilisés, et d'en définir de meilleures conditions d'utilisation.

Annexe 13

Quelques enseignements statistiques de l'évaluation des laboratoires par l'AERES

Hervé Charmettant

Nous nous sommes penchés sur les résultats de deux vagues d'évaluation des laboratoires de recherche en économie et gestion menée par l'AERES (voir la liste en annexe 1) ; 24 laboratoires de la vague A (portant sur le quadriennal 2011-2014 et réalisée en 2009-2010) sont concernés et 27 de la vague D (portant sur le quadriennal 2010-2013 et réalisée en 2008-2009). L'analyse menée ici se fonde sur les données publiques que l'on trouve dans les rapports d'évaluation disponibles sur le site de l'AERES. Les données portent d'abord sur la "production" du laboratoire à travers les indicateurs suivants :

A1 : Nombre de producteurs parmi les chercheurs et enseignants chercheurs référencés en N1 et N2 dans la colonne projet
A2 : Nombre de producteurs parmi les autres personnels référencés en N3, N4 et N5 dans la colonne projet
A3 : Taux de producteurs de l'unité $[A1/(N1+N2)]$
Nombre d'HDR soutenues
Nombre de thèses soutenues
Autre donnée pertinente pour le domaine (à préciser...) PEDR

Elles portent ensuite sur les effectifs selon le tableau suivant. Dans la campagne D, plus ancienne, les effectifs sont présentés pour le seul "bilan". Dans la vague A, apparaissent aussi les effectifs du "projet". Du fait de leur absence dans les rapports des laboratoires évalués lors de la vague D, ces effectifs de "projet" n'ont pas été utilisés :

N1 : Nombre d'enseignants-chercheurs (cf. Formulaire 2.1 dossier de l'unité)
N2 : Nombre de chercheurs des EPST ou EPIC (cf. Formulaire 2.3 du dossier de l'unité)
N3 : Nombre d'autres enseignants-chercheurs et chercheurs (cf. Formulaire 2.2 et 2.4 du dossier de l'unité)
N4 : Nombre d'ingénieurs, techniciens et de personnels administratifs titulaires (cf. Formulaire 2.5 du dossier de l'unité)
N5 : Nombre d'ingénieurs, techniciens et de personnels administratifs non titulaires (cf. Formulaire 2.6 du dossier de l'unité)
N6 : Nombre de doctorants (cf. Formulaire 2.8 du dossier bilan de l'unité et formulaire 2.7 du dossier projet de l'unité)
N7 : Nombre de personnes habilitées à diriger des recherches ou assimilées

Enfin, les résultats de l'évaluation sont présentés de la façon suivante, avec quatre modalités possibles pour chaque rubrique selon un ordre croissant d'appréciation : C, B, A, A+ :

Note de l'unité	Qualité scientifique et production	Rayonnement et attractivité, intégration dans l'environnement	Stratégie, gouvernance et vie du laboratoire	Appréciation du projet
-----------------	------------------------------------	---	--	------------------------

Le but de cette étude était de dégager quelques leçons des statistiques disponibles quant à la politique d'évaluation des laboratoires de recherche en économie et gestion menée par l'AERES. Nous avons essayé de faire ressortir ce qui semblait être les facteurs déterminants des résultats de l'évaluation -synthétisés par la "note de l'unité"- pour fournir des points d'appui à des propositions d'autres formes d'évaluation.

1. La cohérence de la notation

La première investigation concerne la question de la cohérence de la notation ; cohérence d'une part entre les notes des différentes rubriques et la note synthétique, et cohérence d'autre part de la synthèse des notes entre les laboratoires.

Pour le premier point, l'AERES revendique l'importance de la rubrique "qualité scientifique et production" dans la détermination de la "note de l'unité". Cela semble être effectivement le cas car il y a très peu de cas où le classement d'un laboratoire selon l'une ou l'autre note diffère. Seuls cinq laboratoires sont dans ce cas ; 3 laboratoires ayant eu une note globale de B dont la "qualité scientifique et production" est notée A tandis que 2 laboratoires ayant eu A globalement ont une note de B pour cette dernière rubrique.

Pour le second point, il y a des situations où il semble qu'un manque de coordination entre les équipes d'évaluateurs ait eu des conséquences sur la cohérence des évaluations. Par exemple, le cas suivant semble particulièrement incohérent ;

A+	A	A+	A	A
A	A+	A	A+	A

Autre exemple :

A	B	B	A	A
B	B	A	A	B
B	B	A	B	A

Ou encore :

B	C	C	B	B
C	C	B	C	B

Ces exemples, non exhaustifs, peuvent représenter des cas d'arbitraire dans l'évaluation des laboratoires compte tenu des différences d'appréciation des équipes de visite. La coordination

réalisée a posteriori ne réussit pas à contrer ce risque comme ces exemples tendent à le montrer.

2. Les différences d'évaluation selon les vagues

Un second regard a été porté sur les statistiques des résultats des évaluations selon les deux vagues pour lesquelles on dispose des résultats. Le nombre de laboratoires d'économie et gestion est sensiblement le même pour ces deux vagues (27 pour la vague D et 24 pour la vague A). Le tableau suivant présente ces résultats :

	A+	A	B	C
Vague D 2010-2013 (27 labos)	19 %	15 %	52 %	15 %
Vague A 2011-2014 (24 labos)	4 %	50 %	29 %	17 %

Le glissement note B => note A de la vague D à la vague A est très remarquable. Plus de la moitié des laboratoires (52%) ont ainsi la note B pour la vague D. Ce pourcentage baisse considérablement pour la vague A puisqu'il tombe à moins de 30% (29%). Corrélativement, alors que 15% seulement des laboratoires obtiennent la note A pour la vague D, ce pourcentage monte à 50% pour la vague A. Ces chiffres montrent une sensible évolution dans la répartition des notes, dont on ne sait pas si elle tient à des consignes explicites ou à des comportements plus ou moins spontanés des comités de visite.

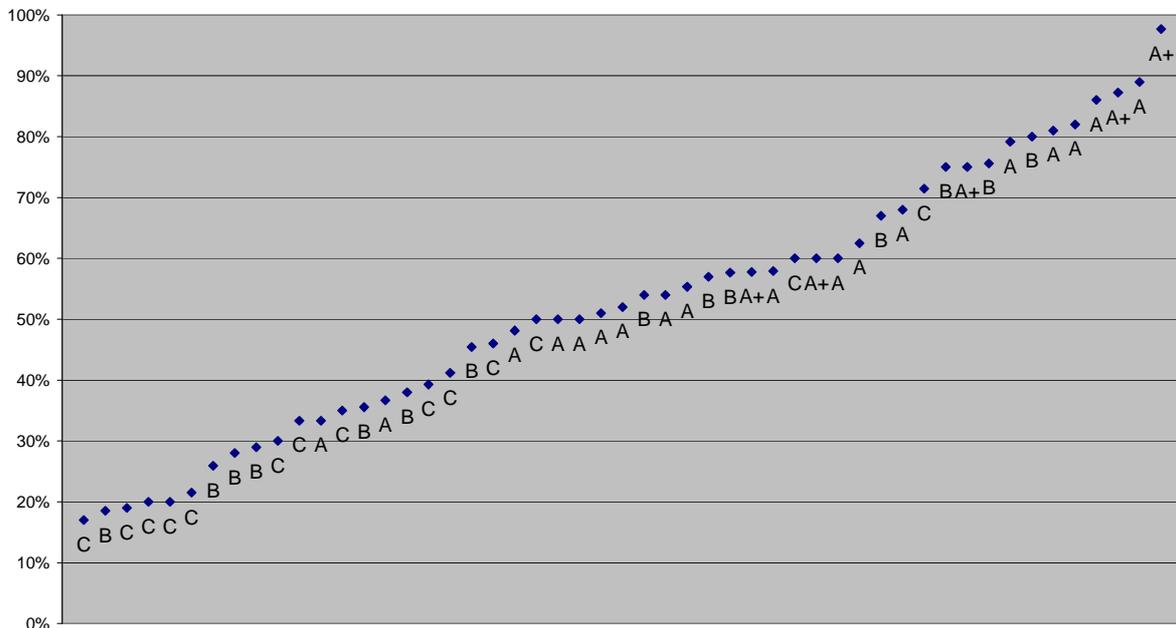
L'économie-gestion avait été pointée du doigt comme étant la discipline où l'évaluation était la moins favorable et il semble qu'une adaptation soit intervenue qui pose un problème d'équité puisque la "note" semble dépendre de **la période** pendant laquelle les laboratoires ont été évalués. Autre constat remarquable, c'est la stabilité de la part des laboratoires notés C, une note-sanction, qui pourrait laisser penser à une pratique de quota ou de "forced ranking" qui pose encore des problèmes d'équité. Enfin, la réduction drastique de la part des A+ (une part divisée par 5 environ) semble indiquer l'aspect d'"excellence" voire d'exception que l'AERES entend donner à cette note.

3. Note de l'unité et "taux de producteurs" :

Le taux de producteurs est réalisé en mettant au numérateur les "producteurs", au sens de l'AERES (soit A1), et au dénominateur la somme des "enseignants-chercheurs" et des "chercheurs" (soit N1 + N2 dans le tableau des effectifs). Il est à noter que ce rapport est déterminé sur une base instable. Tantôt, le rapport (numérateur comme dénominateur) est calculé en référence au bilan, tantôt il est calculé par rapport au projet. Il semble même que, parfois, est considéré au numérateur le nombre de producteurs du **bilan** et au dénominateur le nombre d'enseignants-chercheurs et de chercheurs du **projet**. Il s'agirait alors d'une logique de calcul pour le moins curieuse ne pouvant qu'aboutir à des écarts de résultat très néfastes pour les laboratoires en développement pour lesquels il est prévu que le nombre de chercheurs augmente tandis que le nombre de producteurs est, lui, borné par des effectifs faibles initiaux. Un exemple montre qu'un laboratoire, noté C, s'est vu reconnaître un taux de producteurs de 46 % alors que le calcul, à partir des chiffres du bilan, aurait donné un taux de producteurs de 67 %, ce qui aurait peut-être changé la note de l'unité...

Le graphique ci-dessous permet de montrer que le taux de producteurs n'est pas corrélé très fortement avec les notes des laboratoires, en particulier la note de "qualité scientifique et production" qui devrait avoir un lien fort avec cet indicateur mis en avant par l'AERES :

Note de "qualité scientifique et production" et taux de producteurs



On s'aperçoit de grandes tensions dans ce classement puisque l'on observe des laboratoires notés C, avec des taux de producteurs de près de 70 %, entourés de laboratoires notés A. De même, le dernier laboratoire noté A+ a un taux de producteurs d'à peine plus de 60 % alors qu'un laboratoire noté B a un taux de plus de 80 % ! La qualité des publications (essentiellement le nombre d'articles et la qualité des supports de publication) pourrait certes expliquer ces incongruités statistiques. Mais, nous n'avons pas les moyens de le vérifier puisque ces données ne sont pas publiques et la porte est ouverte à des interprétations en termes d'arbitraire de l'évaluation.

4. La contestation des données

Ce sentiment d'arbitraire peut être confirmé par le fait que les données de producteurs utilisées par l'AERES sont fréquemment contestées par les laboratoires évalués. On en trouve trace dans les courriers de réponse à l'évaluation envoyés par les laboratoires à l'AERES, courriers inclus dans les rapports publics. Cette contestation n'est pas marginale car elle émane du quart des laboratoires de la vague A (6 sur 24) et de près de la moitié de ceux de la vague D (13 sur 27) dont nous rappelons qu'elle a donné des résultats nettement plus sévères. Au total, parmi les 19 laboratoires contestataires, 2 ont pourtant été notés A+, 5 ont été notés A, 8 ont été notés B et 4 ont été notés C. Ce qui montre que la contestation n'émane pas seulement des laboratoires les plus mal notés...plus du tiers des laboratoires contestataires ayant été évalué A ou A+.

Les critiques émanant des laboratoires (droit de réponse) peuvent être résumées ainsi :

- Une contestation de principe, en particulier le fait que l'activité de publication n'est qu'une mesure imparfaite de la production scientifique :

- Le handicap de la pluridisciplinarité : les publications dans les revues non référencées par l'AERES ne sont pas prises en compte
- L'omission des chapitres d'ouvrage, d'articles publiés dans des revues étrangères non anglo-saxonnes, de l'édition d'ouvrages collectifs, de rapports officiels, des articles acceptés mais non encore publiés, les problèmes de date...
- La non prise en compte de facteurs explicatifs tels qu'un congé maternité, des responsabilités dévoreuses de temps, la proximité du recrutement...
- Les problèmes de délimitation des effectifs au dénominateur, en ce qui concerne les ingénieurs détachés ou une politique des laboratoires consistant à s'ouvrir au plus grand nombre, et en particulier, aux non producteurs...

5. Note de l'unité et moyens disponibles

La note de l'unité de recherche a été mise en rapport avec un certain nombre d'indicateurs de moyens disponibles pour la production scientifique ; les effectifs (soit la somme de N1 à N6), le rapport entre "chercheurs" et "enseignants-chercheurs" (soit $N2/N1$), et le nombre de IATOS par "enseignant-chercheur" (soit $N4/N1$). Les résultats sont les suivants :

Note du labo	Effectifs moyens	Rapport chercheurs sur EC	Nombre d'IATOS par EC
C	29	0,02	0,13
B	71	0,19	0,20
A	126	0,30	0,20
A +	161*	0,24	0,28

* si l'on omet un laboratoire aux effectifs exceptionnels, 492 pour le CES de Paris 1, la moyenne des effectifs tombe à 96 pour les laboratoires notés A+.

On peut noter que l'étude établit un lien très fort entre les effectifs de l'unité de recherche et sa notation. Diverses explications peuvent être apportées, certaines allant dans le sens de dire que les laboratoires notés C n'ont pas la "taille critique" et ne bénéficient pas d'effets de synergie comme ce peut être le cas entre des chercheurs plus nombreux. Mais, l'on peut aussi suggérer que les laboratoires de plus petite taille souffrent de handicaps qui n'ont rien à voir avec la qualité de la recherche. D'abord, l'évaluation demande un travail conséquent de collecte des données et de mise en forme que les grands laboratoires peuvent mieux assumer. Autrement dit, la procédure d'évaluation entraîne des coûts rédhibitoires pour ces petits laboratoires. Cette explication est cohérente avec un autre indicateur, celui du rapport des personnels "supports" à la recherche avec le total des enseignants chercheurs, qui montre que les petits laboratoires sont nettement moins bien dotés (seulement 0,13 personnel IATOS par EC).

Le rapport entre les "chercheurs" et les "enseignants chercheurs" est aussi éclairant. La disponibilité des personnels dont le statut est de s'impliquer dans la seule recherche, sans activité d'enseignement, est sans doute très intéressante en termes d'effets d'entraînement et de possibilité de publication (avec des co-auteurs par exemple...).

L'effet-taille semble donc recouvrir des phénomènes qui ne sont pas uniquement liés à la petite taille, mais au fait que les laboratoires de petite taille sont pénalisés par les modalités actuelles d'évaluation et les critères utilisés. On peut cependant penser que ce type de

laboratoire a sa place dans la recherche, en se positionnant sur des objets précis, et que le "small is beautiful" pourrait être observé dans le cadre d'une évaluation moins univoque.

6. Note de l'unité et encadrement de la recherche

Nous avons enfin étudié en quoi la note de l'unité de recherche pouvait être appréciée à partir d'indicateurs de son activité d'encadrement de la recherche, et donc des efforts engagés dans la "transmission". Nous avons ainsi calculé la part des doctorants dans les effectifs (soit $N_6/\text{somme de } N_1 \text{ à } N_6$), le nombre de doctorants par HDR (soit N_6/N_7) et le nombre de thèses soutenues par HDR (soit une donnée de production / N_7).

Note du labo	Part des doctorants	Doctorants par HDR	Nombre de thèses par HDR
C	51 %	4,6	1,8
B	49 %	3,4	1,3
A	44 %	2,6	1,2
A +	50 %	2,9	1,4

La part des doctorants est très faiblement dégressive avec la note, avec une différence de 7 points en moins pour les laboratoires notés A par rapport à ceux notés C (ceux notés A+ ayant par contre une part similaire). Mais, ce qui est beaucoup plus frappant, ce sont les données sur les doctorants ou les thèses soutenues par HDR. La corrélation négative avec la note semble bien établie (avec néanmoins une remontée des indicateurs pour les seuls laboratoires notés A+). On s'aperçoit que le nombre de doctorants par HDR est presque la moitié pour les laboratoires notés A vis-à-vis de ceux notés C. Et le nombre de thèses soutenues par HDR est inférieur d'un tiers, toujours entre les mêmes laboratoires.

On peut assez facilement en déduire que l'engagement dans l'encadrement des jeunes chercheurs peut être pénalisant pour les résultats de l'évaluation des laboratoires telle qu'elle est pratiquée par l'AERES. C'est assez injuste que les chercheurs actuels qui s'activent le plus pour permettre à l'Université de former des futurs chercheurs soient au bout du compte pénalisés en ne pouvant souscrire aux critères de publication qui sont aujourd'hui imposés.

Annexe 14

Pour un débat sur les représentations en économie

G. Colletis

Revue Esprit, décembre 2010

-Extraits-

Nota : toutes les notes de bas de page ainsi que la bibliographie du texte ont été supprimés.

Certains passages ont été soulignés par nous pour attirer l'attention du lecteur et faciliter une lecture rapide.

L'économie se trouve dans une situation critique face aux intérêts qui cherchent à l'instrumentaliser. C'est donc aussi sous cet angle qu'il faut comprendre l'exigence du pluralisme. Celle-ci va au-delà de l'idéal de tolérance propre à toutes les activités intellectuelles. Le pluralisme est une manière pour la communauté des économistes de réfléchir à son rôle social et d'éviter la manipulation. Ce pluralisme, l'AFEP se propose de le décliner selon trois axes.

- Pluralisme des approches conceptuelles au sens où doit être acceptée l'idée que le débat sur ce que sont les concepts de base de l'économie n'est pas clos. À côté du courant néo-classique, existent d'autres traditions de pensée également fécondes, également légitimes, notamment l'institutionnalisme, le keynésianisme, le marxisme, l'évolutionnisme, etc. Cette diversité doit être considérée comme une richesse constitutive de l'économie.
- Pluralisme des points de vue, par quoi nous voulons signifier que l'économie ne se résume pas à la théorie au sens étroit. Il est d'autres manières de faire avancer les connaissances : l'histoire des faits et de la pensée économiques, la réflexion sur la méthodologie, la philosophie économique, la sociologie économique doivent être considérées comme des points de vue légitimes et importants.
- Pluralisme, enfin, des disciplines, car l'économie appartient aux sciences sociales et doit renforcer ses liens avec celles-ci. La situation actuelle qui tend vers une hégémonie extrême de l'économie à leur égard plutôt qu'une collaboration équilibrée est malsaine. La fermeture de l'économie trouve là une de ses racines les plus significatives. Les sciences sociales forment une entité structurée qui doit grandir de manière équilibrée dans ses diverses composantes pour insérer l'économie dans le réel, pour comprendre les faits économiques en tant que faits sociaux.

Selon Postel, Sobel et Tinel, le risque qu'encourt le champ de l'économie actuellement est celui d'une rigidification des canons de scientificité autour d'un paradigme unique et celui, corollaire, d'une stigmatisation académique de tout autre positionnement théorique ou méthodologique. Il est aussi celui d'une fuite en avant de la théorie dans la technicité, afin de se conformer à l'image des sciences de la nature, au risque de couper l'économie des autres sciences sociales et de la priver des richesses de l'interdisciplinarité. Si l'on suit les trois auteurs, alors on peut avancer qu'enfermement idéologique et dérive instrumentale sont les deux maux qui guettent la discipline et sa production académique.

Il a toujours existé, dans la longue histoire des idées, des théories dominantes, des théories dominées, des théories émergentes, des théories progressivement abandonnées, une concurrence, au final, entre différents "paradigmes".

Le problème aujourd'hui n'est pas que le mainstream soit néoclassique ou libéral. Il se situe dans la domination sans partage de ce qui apparaît comme la "pensée unique" autorisée en économie. Le fait que de récents prix Nobel en économie (ou prix assimilés à un Nobel) ait été attribué à des penseurs ne relevant pas de ce courant (Stiglitz, Krugman, Ostrom) ne modifie pas ce constat. Celles qui sont posées comme les grandes revues de la discipline (classées A+), les grandes Universités (notamment américaines), les principales institutions de la discipline sont aujourd'hui tenues par ce courant. A cela, deux explications sans doute. La première est l'appui dont ont bénéficié les représentants de ce courant depuis longtemps et singulièrement depuis la seconde guerre ainsi que les moyens dont ils disposent émanant de différentes institutions ou fondations proches des milieux d'affaires et de certains gouvernements. La seconde explication est le très grand émiettement de ce que l'on désigne parfois comme l'économie "hétérodoxe" constituée de multiples courants.

Sur la base de ce constat –qui mériterait de façon évidente d'être travaillé–, deux postures sont envisageables : tenter de fédérer les courants dits "hétérodoxes" pour disposer d'une doctrine alternative à celle du mainstream, reconsidérer l'économie comme discipline dans son ensemble en mettant en évidence sa polydoxie.

Nous pensons que la première posture n'est ni praticable en raison des différences, voire des divergences paradigmatiques entre courants dits "hétérodoxes", ni souhaitable car une telle entreprise réduirait encore la diversité de la discipline alors que, précisément c'est la réduction de cette diversité qui, aujourd'hui, pose problème.

La seconde posture a notre préférence et correspond aussi à l'orientation que défend l'AFEP lorsqu'elle affirme ou revendique la pluralité des approches conceptuelles et des points de vue. Ajoutons à cela, la pluralité des objets de recherche et des méthodes.

Le mainstream, dès lors que l'on adopte cette posture, apparaît bien comme faisant partie du champ de la discipline, au même titre que les autres courants, fondé sur des présupposés qui lui sont propres, comme le sont les présupposés des autres courants.

La controverse, loin d'être éliminée (ce qui est le cas lorsqu'une croyance l'emporte sur une autre), est revivifiée. La discipline, dans son ensemble, apparaît pour ce qu'elle doit être : polydoxique, c'est-à-dire partagée entre différentes doxas, chacune étant libre d'exprimer sa vision du monde sans être qualifiée comme, alternativement, dogmatique (critique hétérodoxe au mainstream) ou n'entrant pas dans le champ de la discipline (exclusion de ce qui n'est pas mainstream par les tenants de celui-ci).

Une économie polydoxique entre en écho avec la conception qui est défendue ici de l'économiste au service du demos, qui ne produit pas un savoir d'expert sous la forme d'une vérité prétendument dénuée de toute contingence (représentation implicite, intérêts défendus) mais contribue à éclairer les choix possibles et à en inventer de nouveaux.

Sur quelles bases, enfin, évaluer la production scientifique de la discipline ? Telle est la dernière question que nous souhaitons soulever dans ces propos conclusifs.

De nombreux chercheurs ont, dans la dernière période, contesté ce qui s'apparente de plus en plus à une "course aux classements" ou à une "politique du chiffre" telle qu'elle pratiquée désormais (également) dans de nombreux secteurs relevant du secteur non marchand (dans les domaines de la sécurité, de la santé, de l'éducation au sens large).

Cette politique produit incontestablement des effets attendus pour certains mais pervers pour les autres, la majorité sans doute des chercheurs.

Selon A. Caillé, l'expertise anonyme fait presque systématiquement triompher l'opinion moyenne et aseptisée, ce qui "doit être fait dans la discipline". Curieusement, les économistes qui placent au dessus tout les vertus de la concurrence, ici la font disparaître en tentant de graver dans le marbre le paysage des revues tel qu'il est composé par l'élite du moment. Ce système tue tout aussi bien les revues publiées par les sociétés savantes que les ouvrages, comme le montre L. Waters qui évoque à ce propos une "crise du jugement". Pour elle, le livre s'adresse à une intelligence commune alors que les articles cultivent l'ésotérisme, l'hyperspécialisation jargonnante et autoréférentielle.

On peut ainsi se demander si la production académique, du moins en économie, n'est pas en passe de porter atteinte à la production scientifique. Si la production académique est évaluée sur la base quantitative des différents "rankings" (classements des revues, des laboratoires, des Universités) opérés sur des critères qui privilégient "ce qui doit être fait dans la discipline" ou encore la "bonne économie", la production scientifique -entendue comme la tentative de faire avancer la compréhension plutôt que l'explication- se laisse plus difficilement évaluer avec de tels critères. Une évaluation sérieuse de la production scientifique implique une étude non superficielle, par les pairs, de la production réalisée, la confrontation des hypothèses de la recherche avec les termes de la problématique retenue, la cohérence entre ces hypothèses, les concepts choisis et la conduite du projet de recherche, le choix ad hoc des méthodes et outils (sans rejet ni obligation a priori des méthodes formalisées).

Pour l'exprimer en d'autres termes, si le "ranking", de manière générale, apparaît comme le moyen, pour les mieux classés, de renforcer leur domination, de façon plus ciblée, le classement des revues exprime les hiérarchies de la discipline bien davantage que l'innovation scientifique.

De cette perspective, il est parfaitement compréhensible, comme le montre F. Audier, que des chercheurs désignent comme des revues de référence des revues auxquelles ils n'accèdent pas, sauf réseaux particuliers. Parmi les quelque 700 revues classées par la section 37 du CNRS (économie-gestion), seules 8 sont de langue française, seuls 62 articles dépouillés sur 1636 (moins de 4%) sont signés ou co-signés par au moins un Français.

Il est difficile ici de ne pas parler d'hétéronomie en ce sens que les chercheurs français publiants ou cherchant à l'être se contraignent eux-mêmes en endogénéisant des règles de publication (et une langue) produites par et pour d'autres.

Une autre expression de l'hétéronomie est celle des découpages, des rubriques et de leur pondération dans les différents systèmes de classement des revues en économie (JEL, Econlit, CNRS, AERES). Sans développer ici cet aspect, on observera que l'architecture de ces découpages est tout sauf neutre, favorise certains champs de l'économie au détriment d'autres.

Au final, derrière la question de l'évaluation se situent deux enjeux majeurs fortement corrélés : un enjeu général, celui des représentations dominantes -qui cherchent actuellement à être hégémoniques- de la discipline, un enjeu plus opérationnel, celui du contrôle des moyens. Dans ce qui apparaît comme un "marché" de la publication, le versement de primes à la publication, voire "l'achat" de chercheurs publiants, accroissent les performances de publication (bibliométrie) du laboratoire et sa "note" d'évaluation. A son tour, le laboratoire ainsi "noté" obtient des dotations plus fortes, ce qui lui permet de financer les primes, l'achat de chercheurs publiants...

La reproduction élargie de la discipline semble avoir de beaux jours devant elle sauf à considérer que les économistes encore plus nombreux aujourd'hui à relever d'autres approches que celle du mainstream parviennent à se rassembler non pour parler d'une voix (ce qui correspondrait à une nouvelle mais improbable et, surtout, non souhaitable orthodoxie) mais

pour réaffirmer la diversité des représentations, des objets et des méthodes en économie ainsi que la nécessaire insertion de la discipline dans le champ des sciences sociales.

Annexe 15

Une nouvelle catégorie de revues "multidisciplinaires ou d'autres disciplines" (MAD)

La commission 37 a ajouté à sa liste de revues "classées" une longue liste de revues non classées (nommée MAD), qui représentent des revues dans lesquelles les économistes publient ou ont publié.

Compte-rendu de la session d'automne 2010 de la commission 37

-Extrait-

Catégorisation des revues :

La section a une nouvelle fois discuté de la catégorisation des revues. Avant d'achever la mise à jour de la liste (...) la section a mis en place une nouvelle catégorie de revues, dites "multidisciplinaires ou d'autres disciplines" (MAD) dans laquelle elle se refuse à établir un classement.

Remarques :

L'AFEP se félicite particulièrement de la présence dans cette liste de la Revue Française de Sociologie.

La revue Annales, Histoire, Sciences sociales était dans le classement 2007 et 2008 du CNRS dans la catégorie 1 dans le domaine HPE, histoire économique, méthodologie. Dans le contexte actuel de l'évaluation, les enseignants-chercheurs qui ont publié dans cette revue, verront leur situation se dégrader. En effet, au lieu d'avoir une revue de rang 1 (le rang d'une revue est censé refléter la qualité de ladite revue), ils se retrouvent avec une publication qui n'a aucun rang, sur une liste de revues à plat.

Une interrogation : dans le régime actuelle d'évaluation, que vaut une liste à plat à côté d'une liste composée de revues hiérarchisées ?

Catégorie MAD, revues à comité de lecture Multidisciplinaires ou d'Autres Disciplines

Dans le cadre de son travail d'évaluation, la section 37 a identifié ici des supports de publication au delà des champs d'économie et de gestion mais ouverts aux gestionnaires et économistes. La liste comprend donc des revues (francophones ou anglophones) publiant les travaux de chercheurs de nombreuses disciplines dont régulièrement des économistes ou gestionnaires, et des revues de divers champs disciplinaires - démographie, géographie, histoire, mathématiques, philosophie, psychologie, sociologie...- engagées dans un échange avec l'économie ou la gestion.

Actes de la recherche en sciences sociales	<i>social sciences, interdisciplinary</i>
American Journal of Political Science	<i>citizenship, governance, politics</i>
American Journal of Sociology	<i>sociology</i>
American Political Science Review	<i>political science</i>
American Sociological Review	<i>sociology</i>
Annales de l'Institut Henri Poincaré	<i>non-linear analysis</i>
Annales HSS	<i>history, social sciences</i>
Annals of Applied Statistics	<i>statistics, economics</i>
Annual Review of Environment and Resources	<i>environment</i>
Annual Review of Neuroscience	<i>neurosciences</i>
Annual Review of Sociology	<i>sociology, economic sociology</i>
British Journal of Political Science	<i>political science, sociology, economics</i>
Chaos, Solitons and Fractals	<i>non-linear analysis, economics</i>
Comparative Political Studies	<i>comparative politics, economics</i>
Ecological Applications	<i>ecology, policy, economics</i>
Electoral Studies	<i>voting, game theory, economics</i>
Environmental Health Perspectives	<i>environment, health</i>
Ethics	<i>Philosophy</i>
Eurasian Geography and Economics	<i>geography, economics</i>
Evolution and human behavior	<i>interdisciplinary</i>
International Family Planning Perspectives	<i>demography, sociology, economics</i>
International Migration Review	<i>demography, sociology, economics</i>
Journal of American Medical Association	<i>health</i>
Journal of Applied Ecology	<i>ecology</i>
Journal of Applied Psychology	<i>psychology, hrm</i>
Journal of European Public Policy	<i>European public policy, social sciences</i>
Journal of European Social Policy	<i>Social policy</i>
Journal of Experimental Psychology: General	<i>experimental psychology</i>
Journal of Family Psychology	<i>family psychology</i>
Journal of Mathematical Analysis and Applications	<i>applied mathematics</i>
Journal of Nonlinear Science	<i>applied mathematics</i>
Journal of Optimization Theory and Applications	<i>optimization, mathematical economics</i>
Journal of Policy Analysis and Management	<i>social sciences, economics</i>
Journal of Politics	<i>comparative politics, public policy</i>
Journal of Social Issues	<i>social sciences, economics</i>
Journal of the American Statistical Association	<i>statistics</i>
Journal of Theoretical Biology	<i>biological processes</i>
Lancet	<i>medicine</i>
Mathematical Population Studies	<i>demography, economics</i>
Nature	<i>natural sciences</i>
Neuron	<i>neurosciences</i>
New England Journal of Medicine	<i>medicine</i>

Nonlinear Analysis Theory, Methods and Applications	<i>dynamics, applied dynamics</i>
Optimal Control Applications and Methods	<i>optimal control, socio-economics</i>
Philosophy and Public Affairs	<i>Philosophy, political science</i>
Physica D	<i>Nonlinear Phenomena</i>
Political Analysis	<i>political science, empirics</i>
Political Geography	<i>geography</i>
Population Studies	<i>demography</i>
Proceedings of the National Academy of Science	<i>science</i>
Progress in Human Geography	<i>geography</i>
Psychological Review	<i>psychology</i>
Psychological Science	<i>psychology</i>
Public Policy and Administration	<i>public policy</i>
Review of Educational Research	<i>education, psychology, economics</i>
Revue Française de Sociologie	<i>sociology</i>
Science	<i>science</i>
SIAM Journal on Applied Dynamical Systems	<i>engineering, social sciences</i>
SIAM Journal on Applied Mathematics	<i>engineering, finance</i>
SIAM Journal on Control and Optimization	<i>optimization</i>
Sociologie du travail	<i>sociology, hrm</i>
Transactions of the Institute of British Geographers	<i>political geography</i>
Trends in Cognitive Science	<i>Cognitive science, neuroeconomics</i>
Trends in Ecology and Evolution	<i>ecology</i>

Conseils d'usage :

- Pour les chercheurs gestionnaires et économistes, cette liste offre des pistes à ceux qui souhaitent faire connaître leurs travaux au-delà de leurs communautés ou entretenir une démarche d'échanges avec d'autres disciplines

- Pour les évaluateurs, un usage raisonné de cette liste peut aider à caractériser le degré d'ouverture pluri, multi ou interdisciplinaire d'un chercheur et plus sûrement d'une unité ou d'une équipe.