



**Milard Béatrice**

L'évaluation de la recherche par les pairs : les risques d'une formalisation contre performante

Pour citer l'article

Milard Béatrice, « L'évaluation de la recherche par les pairs : les risques d'une formalisation contre performante », dans *revue  $\dot{\iota}$  Interrogations ?*, N°11 - Varia, décembre 2010 [en ligne], <http://www.revue-interrogations.org/L-evaluation-de-la-recherche-par> (Consulté le 27 juillet 2024).

ISSN 1778-3747

---

Tous les textes et documents disponibles sur ce site sont, sauf mention contraire, protégés par la [Licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Pas de Modification 3.0 France](#).



## Résumé

Depuis quelques années, l'évaluation de la recherche en France subit de plus en plus de critiques concernant son mode de fonctionnement et on prétend devoir la réformer en profondeur. Ce texte se donne pour objectif de mieux comprendre les enjeux de l'évaluation de la recherche par les pairs et notamment la question de son organisation, c'est-à-dire ses diverses formalisations. A partir d'entretiens avec des chercheurs, il montre que l'évaluation par les pairs basée sur des textes est un mode de fonctionnement original qui fait de la recherche une « sphère d'échanges » particulière. Les critiques auxquelles ce système est soumis, et qui sont à l'origine des tentatives de réorganisation de l'évaluation scientifique, sont une remise en cause qui risque de détruire les fondements même du fonctionnement collectif de la recherche ainsi que son potentiel créatif.

**Mots clés :** évaluation ; recherche ; pairs ; organisation ; publications

## Summary

For a few years, the evaluation of research in France has undergone more and more criticisms concerning its operating process and one claims to have to reform it in-depth. The objective of this text is to understand the stakes of the peer review evaluation and in particular the question of its organization, i.e. its various formalizations. Starting from interviews with researchers, it shows that peer review based on texts is an original operating process which makes research a particular "sphere of exchanges". Criticisms to which this system is subjected, and which justify the attempts of reorganization of scientific evaluation, are a blamed which is likely to destroy the collective bases of research and its creative potential.

**Keywords :** evaluation ; research ; peers ; organisation ; publications

## Introduction

Depuis quelques années, on a vu émerger en France des discours, parfois même au plus haut niveau de l'Etat, déplorant le manque d'évaluation du secteur de la recherche qui ne serait en définitive soumise à aucun contrôle véritable. Ces propos ont pour beaucoup contribué à produire le tollé et la mobilisation sans précédent que le secteur de la recherche a connu en 2008-2009. Les réactions ont été encore plus véhémentes lorsque l'on a voulu associer à ce soi-disant manque d'évaluation un manque de performance de la recherche française. En effet, la France se positionne d'une manière tout à fait honorable dans l'espace mondial de la recherche scientifique, notamment au niveau des publications et des brevets où elle est généralement classée au 5ème rang mondial [1]. Sans forcément minorer la recherche française, certains critiquent l'évaluation telle qu'elle se pratique - il est vrai de manière singulière - et suggèrent des améliorations de sorte à amener la recherche de ce pays vers un 'niveau d'excellence' : « *La pratique de l'évaluation en France, indispensable dans un système de recherche tourné vers l'excellence, est confiée désormais à l'Agence d'évaluation de la recherche et de l'enseignement supérieur.[...] La troisième ligne directrice de notre politique a consisté à introduire des facteurs de différenciation pour l'attribution des moyens sur le seul critère de l'excellence scientifique, évaluée par les pairs* » [2]. Ces propos laissent entendre, soit qu'il n'y avait pas d'évaluation par les pairs avant la mise en place de cette politique, soit que l'évaluation telle qu'elle se pratiquait n'était pas toujours basée sur l'excellence. Je commencerai donc par montrer que depuis l'avènement de la science moderne il existe des procédures et des dispositifs sur lesquels s'appuie l'évaluation de la recherche qui présente une configuration particulière, peu institutionnalisée donc parfois peu visible pour le politique, et ayant la particularité d'être formalisée par des textes. Je poursuivrai par une analyse des critiques de l'évaluation par les pairs ; en quoi remettent-elles en cause les fondements mêmes du fonctionnement de la recherche ? Je terminerai par la mise en évidence des possibles effets négatifs de l'organisation de l'évaluation autrement que par les dispositifs textuels et les réseaux d'échanges entre pairs.

Dès lors que les institutions publiques (ou privées, avec l'émergence actuelle des fondations) s'attachent à évaluer la recherche, les publications des chercheurs sont considérées comme étant au cœur de cette activité. En effet, même lorsqu'il s'agit d'évaluer un chercheur, une équipe, un laboratoire, voire une université, les publications, auxquels on ajoute parfois, selon les domaines, les brevets et les contrats (qui sont des publications spéciales), se présentent comme le meilleur indicateur. Je baserai donc l'essentiel de mon analyse sur l'étude des publications. Mon propos sera étayé par deux enquêtes en cours. La première est une recherche

sur les publications scientifiques pour laquelle j'ai mené 32 entretiens auprès de chimistes français spécialisés dans des domaines de recherche plutôt académiques à propos de la rédaction, la soumission, l'évaluation et la postérité d'une de leurs publications. La seconde est une enquête sur la visibilité des publications de 72 chercheurs d'un laboratoire de sciences humaines et sociales et destinée à apprécier les nouveaux indicateurs bibliométriques en émergence.

### Une évaluation formalisée par des textes

L'évaluation de la recherche par les pairs est une activité qui remonte à l'émergence de la science moderne, elle-même soutenue par la mise en place des premières revues scientifiques, *Le Journal des sçavans* en France et *Philosophical transactions* en Angleterre [3]. Ce format s'est diffusé assez rapidement parmi les savants du monde européen occidental, notamment parce qu'il était la formalisation ou l'institutionnalisation d'échanges épistolaires antérieurs entre des « cercles savants », académiciens et amateurs de sciences. Dans les faits, le développement de ce mode d'exposition de la recherche a permis une diffusion plus rapide et régulière des résultats scientifiques et la mise en place d'un système d'évaluation nouveau, l'évaluation par les pairs. Comme le raconte David A. Kronick, spécialiste de l'histoire de la presse scientifique, ce système aurait pris, dès le début du XVIIIe siècle, les formes de l'évaluation en aveugle ainsi que l'atteste l'éditorial de la revue *Medical Essays and Observations* de la Royal Society of Edinburgh's, publié en 1731 : « *les mémoires envoyés par correspondance sont distribués d'après leur thématiques aux membres [de la Royal society] qui sont les plus compétents en la matière. Leur identité n'est pas connue de l'auteur* » [4]. En réalité, l'évaluation en aveugle ne se serait vraiment généralisée qu'après la seconde guerre mondiale avec le développement de la « *big science* » [5].

A l'heure actuelle, on peut considérer que l'essentiel des publications produites par les scientifiques paraît dans des revues où existe une évaluation par les pairs. Les pratiques diffèrent selon les disciplines et les pays mais l'exemple de la chimie (très internationalisée) et celui des SHS en France (système plus national) permettront de montrer la spécificité de ce mode d'évaluation.

En chimie, l'évaluation d'une publication est, de manière généralisée depuis l'après-guerre, réalisée à l'aide d'évaluateurs ou de rapporteurs auxquels les éditeurs des revues confient les manuscrits qui leurs sont envoyés. Il s'agit en fait d'un mode de fonctionnement typiquement anglo-saxon auquel les chercheurs français des sciences de la nature ont été confrontés en soumettant, depuis la fin de la seconde guerre mondiale, leurs travaux aux revues américaines. Dès lors, les revues européennes de portée internationale ont progressivement calqué leurs modes d'évaluation sur le modèle américain. Les éditeurs des revues sont des scientifiques très renommés qui sont secondés par dix à quinze éditeurs associés (ou régionaux) auxquels les scientifiques ont le plus directement affaire. Celui-ci confie le texte reçu aux évaluateurs qui produisent un rapport rendant compte de leur évaluation. Ces évaluateurs sont des scientifiques ayant déjà publié dans la revue et choisis par l'éditeur en fonction de leur spécialisation dans le domaine abordé. Ils sont en général au nombre de deux. Si les deux font la même évaluation (refus, acceptation ou révision), cet avis est celui qui est retenu par l'éditeur. S'ils sont en désaccord, l'éditeur s'adresse normalement à un troisième évaluateur, ce qui lui permet d'avoir une opinion prévalente sur le manuscrit.

En SHS, en France, il existe plutôt un fonctionnement sous le mode du comité. La figure de l'éditeur y est bien moins forte puisque bien souvent on a affaire à un comité de rédaction, c'est-à-dire un ensemble de chercheurs chargés de la réalisation des numéros de la revue. La scientificité de la revue est alors affichée par la publication d'un comité scientifique, garant de la légitimité de la revue. Il existe un troisième type de comité, le comité de lecture, qui est un petit groupe de scientifiques dont font parfois partie certains des membres du comité de rédaction. C'est ce groupe qui lit, soit collectivement, soit en désignant ceux d'entre eux qui paraissent les plus proches de la thématique du texte soumis et décide, plus ou moins collectivement, de retenir ou pas l'article soumis.

En réalité les deux procédures ne présentent pas tant de différences. En effet, elles sont soumises à des principes communs. Ainsi, la procédure est toujours la même : les textes sont envoyés par leurs auteurs à des comités ou des conseils qui mettent en place des procédures de sélection de ces textes en confiant l'évaluation à quelques personnes (entre deux et cinq mais jamais plus). De plus, à l'heure actuelle, les différences entre ces deux modes de fonctionnement qui coexistent au sein de la recherche française tendent à se réduire dans la mesure où certaines revues de SHS (parmi d'ailleurs les plus prestigieuses) font le pas de désigner de manière plus large des évaluateurs anonymes (parfois après le rachat par une maison d'édition américaine qui

impose ses modes de fonctionnement à la revue). De même, il ne faut pas totalement associer aux sciences de la nature le premier fonctionnement et aux SHS le second. Comme on l'a dit, certaines revues françaises de SHS ont définitivement adopté le mode de fonctionnement américain. Inversement, il existe des revues américaines très prestigieuses de sciences de la nature qui fonctionnent sur un système de parrainage qui filtre, à la manière d'un comité de lecture tout puissant, les soumissions des auteurs [6].

Une différence notable entre les deux modes de fonctionnement concerne l'anonymat dans le processus d'évaluation. Ainsi, le modèle américain des sciences de la nature fonctionne majoritairement sur le mode du simple aveugle, c'est-à-dire que l'auteur ne sait pas qui a évalué son article mais l'évaluateur connaît le nom de l'équipe qui a soumis le papier. Inversement, dans le mode de fonctionnement des SHS en France, les comités de rédaction garantissent très souvent un anonymat des auteurs évalués alors que le comité de lecture est finalement connu de tous. Les arguments pour justifier telle ou telle pratique sont finalement assez proches : pour les uns, il est inutile de cacher l'identité des auteurs car elle est très simple à déceler ; pour les autres, il importe de la cacher car elle peut avoir une incidence sur l'évaluation. Mais on peut penser ici que ce qui est important n'est pas que l'un ou l'autre soit anonyme, mais qu'il y ait au moins une des parties qui le soit de sorte que l'on ne soit pas dans la situation d'échange interpersonnel, ni dans une situation de « marché » anonyme mais dans un « entre-deux » qui fait, en grande partie, la particularité de ce secteur. Cet entre-deux est totalement organisé par un système d'échange de textes indispensable à son fonctionnement.

Plus généralement, les deux modes d'évaluation ont également en commun de compiler un ensemble d'évaluations qui sont produites avant la parution du texte, tout simplement parce qu'ils filtrent les publications en fonction des qualifications des auteurs, leurs origines institutionnelles, en fonction de leur appartenance à la sphère d'activité scientifique, mais aussi en fonction des propos qui sont tenus. Ainsi, chaque publication est la compilation d'évaluations en amont et en aval de sa parution. Si l'on prend l'exemple d'une publication ordinaire, parue dans une revue « moyenne », par un chercheur universitaire français en milieu de carrière, il y a plus d'une cinquantaine de personnes à « convaincre » pour que ce résultat paraisse ; trois évaluateurs pour le master, cinq pour la thèse, vingt pour la qualification auprès du CNU, quinze pour la commission de spécialistes ou le comité de sélection, quinze pour le comité de lecture ou *editorial boards* de la revue, trois pour les évaluateurs désignés auxquels il faudrait ajouter les trois collègues qui relisent le texte avant qu'il soit soumis et bien sûr l'éventuelle équipe (techniciens, administratifs et autres chercheurs) qu'il a fallu mobiliser pour la réalisation de la publication [7]. Cette compilation d'évaluations, toujours réalisée par l'entremise de textes (mémoires, thèses, publications antérieures...), est assez spécifique au domaine et parfois mal cernée par les médias ou les politiques. Prenons l'exemple d'une journaliste qui profitant du mouvement des enseignants-chercheurs contre la réforme des universités pour réaliser un article sur cette question s'était « renseignée » et, après avoir affirmé qu'un « quart des enseignants-chercheurs ne publie pas », avait ajouté que « [en plus] est considéré comme « publiant » un enseignant qui a satisfait à un nombre minimal d'articles dans des revues reconnues, en général un à trois en quatre ans. Dans le secteur des sciences de la vie, alors que trois publications dans des journaux d'impact moyen suffisent, une publication dans *Nature* ou *Science* permet de cataloguer l'enseignant comme publiant. » [8]. On comprend ici que le système n'est pas toujours bien connu du profane et donc pas compris...

En aval de la publication, les évaluations réalisées lors de la soumission à une revue vont avoir des répercussions à différents niveaux de la recherche. Elles seront en effet impliquées sous la forme de citations que la publication va recevoir au sein d'autres textes mais aussi dans les différents documents qui contribuent à évaluer les chercheurs, les équipes, les laboratoires et même les universités et autres organismes de recherche ou tout simplement à produire d'autres documents (d'autres publications mais aussi les réponses à appel d'offre, les demandes de financement, etc.).

Pour décrire ce mouvement, Latour parle de « cycle de crédibilité » [9]. Mais, en fait de cercle, on pourrait utiliser l'image de l'entonnoir à deux bouts, le moment le plus resserré correspondant aux évaluations par les pairs, quel qu'en soit le mode. A ce titre, elles contribuent en grande partie à réaliser l'autonomie du « champ scientifique » [10], à définir la « discipline » [11], la « sphère d'échanges » [12] de ce secteur d'activité. En effet, du fait de la compilation de textes qui caractérise leur mode de fonctionnement, elles sont un « nœud », un moment de formalisation qui a des implications à plusieurs niveaux de l'activité passée et future des individus et des institutions. L'entremise des textes au sein de ces échanges permet un fonctionnement relationnel assez singulier entre les chercheurs impliqués (*les pairs*) qui n'est pas de l'institutionnalisation de l'évaluation, comme on l'avait interprété peut-être un peu hâtivement [13], ni un « marché » parfait (l'anonymat est rarement total comme le voudraient parfois un peu idéalement les scientifiques eux-mêmes), ni le fruit d'une relation totalement interpersonnelle puisque, du fait de l'anonymat, la relation n'est jamais

symétrique. Quelles sont les critiques à l'encontre de l'évaluation par les pairs et que permettent-elles de comprendre de ce fonctionnement ?

### Les mises en cause de l'évaluation par les pairs

Certains prétendent que même si la recherche est évaluée, elle l'est mal. Plusieurs arguments sont développés et parfois même venant des chercheurs eux-mêmes. Ainsi, une grande enquête sur les pratiques et représentations du *peer review* a été publiée en 2008 [14], suite notamment à la mise en cause en 2002 par une revue américaine réputée, le *Journal of American Medical Association*, du système de *peer review*, soulignant que l'on ne savait rien en définitive de « l'efficacité » de cette procédure [15]. Le questionnaire a réuni les réponses de plus de 3000 scientifiques de disciplines diverses au travers le monde, tant sur leurs expériences d'auteur que sur celles d'évaluateur et d'éditeur, à partir d'un panel de chercheurs ayant récemment publiés dans des revues recensées par le *Science Citation Index*, donc sur le modèle américain d'évaluation des publications.

Les critiques à l'encontre du système de l'évaluation par les pairs renvoient essentiellement à sa mise en œuvre : il est jugé trop lent par les auteurs et chronophage du point de vue des évaluateurs. Mais d'une manière générale, il est apprécié par les personnes interrogées qui le jugent indispensable (93%) et performant (90%). Cependant, l'enquête souligne la pratique générale de l'évaluation en simple aveugle (56%) alors que les chercheurs plébiscitent le double aveugle pour son « objectivité » ainsi que l'*open review* et le *post review* pour leur caractère « démocratique ». Il y a donc, chez les chercheurs, un hiatus entre une pratique de l'évaluation par les pairs et l'idéal de cette pratique. C'est bien souvent la prévention contre une critique de 'copinage' qui provoque ces prises de position [16]. Il existe effectivement des affinités, des « petits arrangements » qui viennent perturber le cours « objectif » de l'évaluation ainsi que le rapporte spontanément cette chimiste : « *on a droit à trois figures mais on a triché un peu... [en effet, il y a quatre figures] ... parce que l'éditeur, il aime bien X [le directeur de l'équipe] et quand ça vient de l'équipe, il y a toujours plein de photos...* » (Nadège [17], chimiste). Cependant l'importance de cet aspect dans le fonctionnement global de la recherche est assez faible (l'exemple concerne une figure en plus dans la publication). Il faudrait en effet qu'un chercheur ait un capital social démesuré pour parvenir à maintenir sur le long terme des relations permettant d'outrepasser régulièrement, et à tous les niveaux décrits ci-dessus, les évaluations de son travail.

Sans parler de « copinage » et de ses effets 'décrédibilisateurs', il existe bien des relations d'interconnaissance avec son lot de sympathie et d'antipathie entre les chercheurs, tel que le rapporte ce chimiste ; « *j'ai reçu ce matin, le manuscrit de quelqu'un que je connais très bien. Je l'ai déjà invité à la maison. Je le connais vraiment très bien ! Je vais faire la correction comme si c'était quelqu'un que je ne connais pas. Eventuellement je vais lui dire, j'ai eu ton article à arbitrer... Bien évidemment je ne suis pas censé le faire !* » (Ronan, chimiste). Un autre chercheur raconte qu'il a été informé par l'évaluateur (censé rester anonyme) du fait qu'il avait accepté son article : « *C'est un grand physicien que je connais, il m'a téléphoné pour me dire qu'il avait accepté mon article* » (Henri, chimiste). Peut-on considérer que ces relations interpersonnelles sont particulièrement rédhibitoires pour le fonctionnement de l'évaluation par les pairs ? En réalité, comme cela a été montré dans d'autres domaines, la sociabilité entre les acteurs est non seulement souhaitable mais nécessaire au fonctionnement de cette sphère d'échange. Or, la soumission et l'évaluation des articles sont justement des moments où s'entretient la sociabilité professionnelle des chercheurs [18]. Le choix des références qui vont figurer dans l'article en est un révélateur : « *Dans un journal européen, on a plus tendance à mettre des auteurs européens, parce qu'on sait que les éditeurs envoient les textes aux auteurs dont ils voient les noms dans les textes qui sont aussi plutôt de la nationalité du sujet* » (Sophie, chimiste). L'activation de cette sociabilité a cela de particulier qu'elle intrique complètement les relations intellectuelles et sociales : « *quand l'éditeur reçoit un article, il regarde les citations, s'il y a les miennes, vous pouvez être sûr que ça m'arrive...* » (Jean-Paul, chimiste), ce qui peut parfois entraîner des déconvenues : « *C'est Y. [le directeur de l'équipe] qui a reconnu X. [l'évaluateur]... On pense qu'il a été vexé parce qu'on ne le citait pas beaucoup, c'est ça qui l'a froissé un peu...* » (Cécile, chimiste).

Certains pensent que la recherche est mal évaluée parce que les pairs qui évaluent les publications et les auteurs sont redondants, finalement un tout petit monde, replié sur lui-même. En effet, dans un domaine de recherche donné, les spécialistes ne sont pas si nombreux, ce qui explique d'ailleurs qu'ils/elles finissent par se connaître personnellement. Cependant, comme je l'ai dit plus haut, que ce soit les collaborateurs, les programmes, les contrats, le petit monde est finalement bien vaste dans la mesure où il peut même franchir les frontières du monde scientifique avec des financements privés ou d'autres organismes publiques. Par ailleurs,

chaque chercheur n'appartient par un petit monde étanche. Dans l'étude des relations personnelles entre les 15 chercheurs interrogés et les 2381 auteurs cités en références dans leur article, seuls 10% leur sont très proches, 25% sont des collègues bien connus et 15% des collègues bien identifiés mais avec lesquels ils n'ont pas d'interaction. Il reste que 50% des noms qu'ils citent sont des personnes qui leur sont inconnues, la plupart du temps pour des raisons de distance géographique ou de spécialisation disciplinaire, mais aussi parce que ce sont des jeunes chercheurs, doctorants ou post-doctorants. La recherche n'est donc pas un seul petit monde mais une « *mosaïque de petits monde qui se touchent sans s'interpénétrer* » [19] et l'activité d'évaluation par les pairs est un moment essentiel de la construction de ces petits mondes. On peut regretter que le *peer review* n'ouvre pas assez à l'interdisciplinarité [20] mais on peut aussi considérer que « *cela donne aux individus la possibilité de passer [...] d'un milieu moral à un autre et encourage cette expérience fascinante, mais dangereuse qui consiste à vivre dans plusieurs mondes différents, certes contigus, mais par ailleurs bien distincts.* » [21].

La recherche est parfois jugée comme étant mal évaluée parce que la science étant, au même titre que d'autres secteurs, un système stratifié et hiérarchisé, le système de publication aurait tendance à reproduire ces inégalités. Ainsi, « *l'effet Mathieu* » [22] et « *l'effet Mathilda* » [23] montrent qu'un scientifique ayant déjà la reconnaissance aura, à travail égal, plus de reconnaissance qu'un/e autre, moins connu/e. Ces discriminations sont bien certaines et, comme tous les autres secteurs de la vie sociale, la recherche est loin d'être exempte d'injustice et de ségrégation, non plus que le système de publication, où l'on a souligné le poids important des « *gate keepers* » [24]. Il existe néanmoins une spécificité de la recherche qui garantit, si ce n'est une complète horizontalité des relations, une moins forte ségrégation, et qui s'organise notamment lors du processus d'évaluation de la recherche : sa relative autonomie et, surtout, l'interchangeabilité des « rôles ». Ainsi dans l'enquête quantitative internationale [25], 90% des chercheurs qui avaient publié dans une revue recensée par le SCl avaient également évalué des textes pour ces revues. Ils peuvent, par ailleurs, être ou espérer devenir éditeur associé pour un journal de prestige : « *on n'est pas éditeur associé toute sa vie, seulement pendant quelques années. Je ne l'ai pas encore été, mais j'aimerais l'être...* » (Gilbert, chimiste). Ils sont donc tous, en réalité ou en puissance, juges et parties, ce qui finalement, sur des objets de recherche assez ténus et donc des collectifs assez réduits, produit des formes de régulation des pratiques. L'autonomie est, à ce titre, essentielle. La situation particulière que représente l'évaluation par les pairs du financement de la recherche (*funding peer review*) en est une preuve. Ce mode de fonctionnement (qui se généralise en France depuis la mise en place de l'ANR) a été qualifié par les sociologues des sciences de particulièrement paradoxal : les évaluateurs sont en effet mobilisés par les politiques pour préserver le caractère auto-régulateur de la profession et, en même temps, sont censés ouvrir la recherche sur le monde social [26]. Il y a vraiment un « *hiatus* » entre les normes mertonniennes (l'éthos scientifique) et les évaluations demandées par l'institution (régulation) [27]. En réalité, le fait même d'être, dans une certaine mesure, 'centré' sur lui-même garantit au monde scientifique une certaine horizontalité des relations.

Que faut-il pour que les « copinages », les « petits mondes », les « effets Matthieu », le poids des « *gatekeepers* » et les autres sources d'inégalités dans les évaluations ne soient pas trop discriminants ? Il faut que le système de publication connaisse un niveau de développement élevé, avec des ramifications un peu indépendantes les unes des autres de sorte à amoindrir (en les multipliant et en les rendant ainsi chacun moins fort) les points de pouvoir, de centralité et donc de domination et d'inégalité. Vers quoi tendent les transformations les plus récentes du système d'évaluation de la recherche en France ?

### **Les effets préjudiciables de la formalisation de l'évaluation**

Il y aurait actuellement urgence à réformer le système d'évaluation de la recherche en évaluant 'plus' et donc 'mieux', en formalisant le système de recherche et d'innovation grâce à une « *stratégie nationale* » [28] ... Or, je voudrais montrer dans cette dernière partie qu'il est au contraire néfaste de vouloir 'organiser' (dans le sens premier du terme) l'évaluation de la recherche.

Des indicateurs bibliométriques individuels sont mis au point, destinés à mesurer la performance individuelle des chercheurs. Depuis longtemps, on calcule, à partir de bases de données bibliographiques telles que le *Web of Science* de l'entreprise Thomson Reuters (anciennement le *Science Citation Index*, le *Social Science Citation Index* et le *Art and Humanities Index* de l'Institut for Scientific Information), la « visibilité » des revues, « l'*impact factor* », c'est-à-dire le nombre moyen de citations qu'une revue a occasionné depuis un certain nombre d'années, en général trois ans. Rapporté aux performances individuelles, le facteur d'impact prend souvent la forme du *h-index* (ou facteur *h*), c'est-à-dire le nombre *h* de publications ayant au moins *h* citations. Plus le *h*



-index est élevé, plus le chercheur est considéré comme visible. Or, le recours aux mesures et classements individuels présente plusieurs effets pervers.

Le premier est le plus fondamental. Il consiste à laisser entendre que l'évaluation par les pairs n'est pas suffisante pour garantir l'importance d'une connaissance scientifique et qu'il faut recourir à des indicateurs de visibilité. C'est donc bien une mise en cause de ce système.

Le deuxième effet pervers est technique : l'application des outils bibliométriques aux auteurs individuels se fait souvent au mépris de l'impératif statistique de base consistant à traiter des agrégats de données suffisamment larges pour « faire jouer le hasard » et non des données sur une seule personne pour lesquelles le contexte individuel et social est trop prégnant [29]. L'enquête que nous avons menée récemment a consisté à tester empiriquement les indicateurs bibliométriques issues de plusieurs bases disponibles (Web of science (WOS), Francis, Google, Google Scholar et Harzing) concernant 72 chercheurs d'un laboratoire pluridisciplinaire de SHS [30]. Concernant les citations, le WOS et le *h*-index de Harzing ont tendance à ne pas mettre en évidence les mêmes auteurs. En réalité, on est face ici à deux formes très distinctes de visibilités. Une visibilité au travers de la littérature académique, celle qui est repérée par le WOS (les 3000 revues ainsi que certains des articles « sélectionnés en fonction de leur pertinence » dans 10 000 revues par l'Institut of Scientific Information). L'autre visibilité correspondant à Harzing est complètement dépendante de l'outil Internet : les citations comptées par Harzing correspondent à des liens, donc à des pages que certains (les auteurs eux-mêmes ou bien d'autres) ont eu la démarche de mettre sur le Web et qui ont été reconnues par Google Scholar comme « académiques ». Par contre, ces citations présentent l'avantage de mettre en avant un certain nombre d'auteurs qui ne le seraient pas par les bases plus traditionnelles. Rappelons également que le *h*-Index et le nombre de citations par référence ne sont pas significativement dépendants du sexe, de l'âge et du statut, contrairement aux citations repérées par le WOS qui présentent toutes les caractéristiques des « avantages cumulatifs ». On a donc affaire, avec les indicateurs « citations » de Harzing, à un outil qui se présente, dans les résultats, comme plus démocratique. Mais le restera-t-il longtemps si on lui accorde le pouvoir d'évaluer les chercheurs ?

Le troisième effet pervers est d'ordre culturel. En effet, les disciplines ou spécialités scientifiques ont des histoires différentes et leurs pratiques le sont aussi. Les formes de production sont parfois spécifiques et le recours à la citation n'est pas toujours envisagé de la même façon. On sait qu'une spécialité naît, vit, s'affaiblit et parfois meurt et que le système relationnel (dont les citations) est fortement lié au degré d'académisation de la spécialité [31]. Quand l'outil bibliométrique est envisagé à des fins d'analyses (avec une vraie maîtrise de la statistique afférente), il peut justement montrer les dimensions culturelles des disciplines et des spécialités. Par contre, il est délicat d'avoir un même outil d'évaluation pour des pratiques très différentes. Une enquête un peu similaire à la nôtre menée aux Etats-Unis en 2006 a comparé les citations entre le WOS et Google Scholar. Elle a montré, elle aussi, l'extrême diversité des pratiques selon les disciplines de sciences sociales et de sciences de la nature [32].

Mais l'effet peut-être le plus pervers est pragmatique et social. En effet, ce classement entraîne des pratiques intempestives d'auto-citation et de signatures d'articles qui ne sont probablement pas généralisables mais qui montrent que l'outil bibliométrique, dès lors qu'il est utilisé à des fins évaluatives et personnalisées, va performer l'activité, c'est-à-dire que, progressivement, les règles de fonctionnement se fondent dans la logique de l'outil (comme la réalité est dans son expression). Basé sur ces critères, formatant en partie et progressivement les pratiques et la visibilité, le système d'évaluation est dorénavant un enjeu essentiel. A ce titre, l'interface Google Scholar et son émanation Harzing sont symptomatiques de cette tendance. Ils sont typiquement des outils de visibilité émanant directement de démarches personnelles : la mise en ligne de certains documents et le signalement de leur caractère académique à Google Scholar, que ce soit par les éditeurs, les bibliothèques ou la maîtrise des méta-moteurs utilisés par Google Scholar. Cette visibilité, que l'on peut renforcer par certaines pratiques dont les règles sont formalisées, va jusqu'à entraîner des phénomènes de « *game playing* » chez certains auteurs, c'est-à-dire que finalement la recherche est vue comme un jeu au cours duquel il s'agit d'avoir les résultats les plus performants...

Les mesures et les classements bibliométriques des revues ont eux aussi entraîné certaines évolutions, plus économiques celles-là. En fait, les taux de citation des revues sont monnayés par les firmes éditoriales commerciales qui progressivement sont passées de diffuseurs à propriétaires des revues. Plus le facteur d'impact de la revue est important, plus cher sera l'abonnement payé par les bibliothèques et les organismes qui achètent les revues. On peut penser que ces facteurs d'impact sont à l'origine et organisent en grande partie le marché qui s'est progressivement constitué autour de la publication. Cela entraîne un phénomène bien connu des analystes de l'information scientifique, à savoir qu'en définitive, les organismes publics

achètent deux fois les mêmes informations à ces firmes : ils paient les abonnements aux revues des bibliothèques ou des organismes pour que les lecteurs (des chercheurs) accèdent à des résultats produits par des chercheurs financés en grande partie avec de l'argent public ! A présent, certaines revues demandent même des sommes d'argent aux organismes de recherche pour soumettre les articles des chercheurs, faisant payer une troisième fois l'information scientifique en question. La constitution d'un marché n'est pas incompatible avec le fonctionnement en réseau, sauf s'il est concentré entre les mains de quelques firmes, ce qui est de plus en plus le cas des maisons d'éditions scientifiques. On peut considérer que cela n'a pas d'importance si les contenus, eux, ne sont pas contrôlés. Mais dès lors qu'il n'y a que quelques personnes à convaincre pour ne pas publier un résultat, on court un risque de censure (par exemple) bien plus fort que si un nombre plus important de personnes, organisées en réseau de surcroît, sont à même de rendre un résultat visible.

Les revues sont donc mesurées, classées et cela a des conséquences économiques qui peuvent avoir des implications scientifiques. Mais la tendance à la classification entraîne d'autres effets, encore une fois plus pragmatiques. Prenons l'exemple de la classification des revues de sciences humaines et sociales en A, B et C proposée par l'AERES (il semble que ce classement soit sur le point d'être abandonné, mais le fait d'être ou non classé AERES présente le même problème). Sans même parler de la difficulté de départager sur un mode consensuel toutes les revues, un tel classement risque d'entraîner le fait que les revues classées en A vont concentrer la plupart des soumissions d'articles et leur gestion va s'en trouver d'autant compliquée et que, inversement, les revues non cotées émergeront encore plus difficilement avec ce classement, qui aura donc tendance à figer les supports d'édition.

Le problème est un peu similaire concernant les experts qui semblent être le mode choisi par l'AERES pour procéder aux différentes évaluations qui lui incombent. Les experts sont forcément choisis pour leur déjà grande production et visibilité. Or, la production de la communauté scientifique correspond à la « loi de Lotka », qui stipule que beaucoup de scientifiques produisent peu et peu produisent beaucoup. Le vivier des experts est donc assez faible et, très vite, ils vont se trouver submergés par la masse des expertises à produire aux différents niveaux.

La concentration des revues et des experts est directement liée à la volonté politique de formaliser le système d'évaluation de la recherche. Elle présente une toute autre forme d'organisation que l'évaluation par les pairs qui, même si elle n'est pas exempte de dysfonctionnements, pose l'horizontalité comme principe de base. Au mieux l'évaluation 'organisée' (c'est-à-dire avec des organes spécifiques) risque de ne servir à rien, par manque de temps et donc de sérieux ou parce qu'elle sera finalement redondante avec l'évaluation effectuée par les pairs lors des soumissions ; au pire elle va entraîner un dévoiement du système en provoquant de la surenchère de publications (de ce fait plus forcément très pertinentes) et en créant des hiérarchies fortes et renforçant des pouvoirs très localisés.

## **Conclusion**

L'évaluation par les pairs fonctionne sur la base d'échanges de textes et sur le modèle du réseau, ou plus exactement de la chaîne de médiation. Ainsi, un auteur est évalué via un texte par un autre chercheur qui a été évalué par un autre, qui lui-même l'a été par un autre encore (ou bien même par le premier) etc., sans que la chaîne relationnelle ne soit vraiment connue. En effet, même s'il n'y a parfois pas d'anonymat ou si celui-ci est dévoilé lors de l'évaluation, le fait de fonctionner par paire, une publication - un évaluateur et entre pairs garantit un chaînage extrêmement développé du système d'échanges d'évaluations que personne ne connaît totalement et surtout ne maîtrise vraiment (même les éditeurs, à condition qu'ils soient nombreux et changeants sur la scène scientifique). C'est cette impossible maîtrise qui, en quelque sorte, produit du contrôle social.

L'horizontalité et le caractère non maîtrisé du système d'échanges concernant les évaluations sont aussi des potentiels générateurs « d'innovations ». En effet, des échanges inédits peuvent se produire, des chaînes relationnelles être activées, des cliques se former à l'occasion du processus d'évaluation par les pairs. Une évaluation plus formelle, encadrée dans de l'organisationnel, risque de rendre cette activité plus mécanique, trop bureaucratique et donc moins créative.



## Bibliographie

- Bourdieu Pierre, « Le champ scientifique », *Actes de la recherche en sciences sociales*, n° 2-3, 1976, p. 88-104.
- Chubin Daryl E. et Hackett Edward J., *Peerless Science : Peer Review and US Science Policy*, State University of New York Press, New York, 1990.
- Crane Diana, « The gatekeepers of sciences : some factors affecting the selection of articles of scientific journals », *American sociologist*, vol. 2, n°1, 1967, p. 195-201.
- Filliartreau Ghislaine, *Rapport de l'OST. Indicateurs de sciences et de technologie*, Paris, Economica & OST, 2008.
- Gingras Yves, « Du mauvais usage de faux indicateurs », *Revue d'Histoire Moderne et Contemporaine*, vol 5, n° 55-4bis, 2008, p. 67-79.
- Grossetti Michel, *Sociologie de l'imprévisible. Dynamiques de l'activité et des formes sociales*, Sociologie d'aujourd'hui (Coll.), Paris, Presses Universitaires de France, 2004.
- Jasanoff Sheila, « Peer Review in the Regulatory Process », *Science Technology Human Values*, vol. 10, n° 3, 1985, p. 20-32.
- Jefferson Tom, Wager Elizabeth et Davidoff Frank, « Measuring the Quality of Editorial Peer Review », *JAMA*, 287, n°21, 2002, p. 2786-2790.
- Kousha Kayvan et Thelwall Mike, « Google Scholar citations and Google Web/URL citations : A multi-discipline exploratory analysis », *Journal of the American Society for Information Science & Technology*, n°58, 2007, p. 1055-1065.
- Kronick David A., *A history of scientific and technical periodicals : the origin and development of scientific and technologic press. 1655-1790*, New York, Scarecrow Press, 1962.
- Latour Bruno et Woolgar Steeve, *La vie de laboratoire. La production des faits scientifiques* [1979], Paris, La découverte, 1988.
- Mallard Grégoire, Lamont Michèle et Guetzkow Joshua, « Fairness as Appropriateness : Negotiating Epistemological Differences in Peer Review », *Science Technology Human Values*, vol. 34, 2009, p. 573-606.
- Merton Robert K., « The Matthew Effect in Science : The reward and communication systems of science are considered », *Science*, 159, n°3810, 1968, p. 56-63.
- Milard Béatrice, « La territorialisation de l'activité scientifique vue à travers les publications » in *La territorialisation de l'enseignement supérieur et de la recherche en Europe. France, Espagne, Portugal*, Grossetti Michel et Losego Philippe (dir.), Paris, L'Harmattan, 2003, p. 109-194
- Milard Béatrice, « La mise en forme des publications scientifiques : entre routines, contraintes et organisation de l'expérience collective », in *Sociologie des arts, sociologie des sciences*, Gaudez Florent (dir.), L'Harmattan, Paris, 2007, p. 203-212.
- Milard Béatrice, « La soumission d'un manuscrit à une revue : quelle place dans l'activité scientifique des chercheurs ? » *Schedae (Presses Universitaires de Caen)*, n°1, 2008, p. 1-12.
- Mullins Nicolas C., « The development of a scientific specialty : the phage group and the origin of molecular biology », *Minerva*, vol. 10, n° 1, 1972, p. 51-82.
- Park Robert, « La ville. Propositions de recherche sur le comportement humain en milieu urbain » [1925] in *L'école de Chicago. Naissance de l'écologie urbaine*, Grafmeyer Yves et Joseph Isaac, Paris, Aubier, 1984, p. 83-130, p. 121.
- Price Derek, *Little science, big science*. New York, Columbia University Press, 1963.

- Rennie Drummond, « Editorial peer review : its development and rationale in Peer Review in Health Sciences », Godlee F. et Jefferson T. (dir.), London, BMJ Books, 2002, <http://resources.bmj.com/bmj/pdfs/r...> (page consultée le 22 mai 2010).
- Rossiter Margaret W., « The Matthew Matilda Effect in Science », *Social Studies of Science*, vol. 23, n°2, 1993, p. 325-341.
- Travis G.D.L. et Collins Harry M., « New Light on Old Boys : Cognitive and Institutional Particularism in the Peer Review System » *Science, Technology, & Human Values*, vol. 16, n° 3, 1991, p. 322-341.
- Ware Mark et Monkman Mike, *Peer review in scholarly journals : perspective of the scholarly community - an international study*, Londres, Publishing Research Consortium, 2008.
- White Harrison, *Identity and Control*, Princeton, Princeton University Press, 1992.
- Zuckerman Harriet et Merton Robert K., « Patterns of evaluation in science - Institutionalisation, structure and functions of referee system », *Minerva*, vol. 9, n°1, 1971, p. 66-100.

---

## Notes

- [1] G. Filliartreau (dir.), *Rapport de l'OST. Indicateurs de sciences et de technologie*, Paris, Economica & OST, 2008 ; cf. aussi B. Milard, « La territorialisation de l'activité scientifique vue à travers les publications » in *La territorialisation de l'enseignement supérieur et de la recherche en Europe. France, Espagne, Portugal*, M. Grossetti et P. Losego (dir.), Paris, L'Harmattan, 2003, p. 109-194.
- [2] Discours M. Goulard, AERES, décembre 2007. <http://www.enseignementsup-recherch&hellip> ; (page consultée le 15 mai 2010)
- [3] D.A. Kronick, *A history of scientific and technical periodicals : the origin and development of scientific and technologic press. 1655-1790*, New York, Scarecrow Press, 1962.
- [4] DA. Kronick, « Peer review in 18th-century scientific journalism », *JAMA*, n° 263, 1990, p. 1321-1322, cité par D. Rennie, « Editorial peer review : its development and rationale in Peer Review in Health Sciences », F. Godlee et T. Jefferson (dir.), Second Edition, London, BMJ Books, 2002, <http://resources.bmj.com/bmj/pdfs/r&hellip> ; (page consultée le 15 mai 2010), p.2 [citation traduite par nos soins].
- [5] Price, D. S., *Little science, big science*, New York, Columbia University Press, 1963.
- [6] Citons par exemple la revue *Proceedings of the National Academy of Science* qui exige pour la soumission d'un article que les auteurs puissent être recommandés par trois membres de l'*Editorial Board* de la revue et trois membres de la National Academy of Science, cf. <http://www.pnas.org/site/misc/iforc&hellip> ; (page consultée le 22 octobre 2010).
- [7] B. Milard, « La mise en forme des publications scientifiques : entre routines, contraintes et organisation de l'expérience collective », in *Sociologie des arts, sociologie des sciences*, F. Gaudez (dir.), L'Harmattan, Paris, 2007, p. 203-212.
- [8] M.-E. Pech, « Un quart des enseignants-chercheurs ne publient pas. », *Lefigaro.fr*. 12/02/2009. <http://www.lefigaro.fr/actualite-fr&hellip> ; (page consultée le 15 mai 2010)
- [9] B. Latour et S. Woolgar, *La vie de laboratoire. La production des faits scientifiques*, [1979], Paris, La découverte, 1988.
- [10] P. Bourdieu, « Le champ scientifique », *Actes de la recherche en sciences sociales*, n° 2-3, 1976, p. 88-104.
- [11] H. White, *Identity and Control*, Princeton, Princeton University Press, 1992.
- [12] M. Grossetti, *Sociologie de l'imprévisible. Dynamiques de l'activité et des formes sociales*, Sociologie d'aujourd'hui (Coll.), Paris, Presses Universitaires de France, 2004.
- [13] H. Zuckerman et R. K. Merton, « Patterns of evaluation in science - Institutionalisation, structure and functions of referee system », *Minerva*, vol. 9, n°1, 1971, p. 66-100.
- [14] M. Ware et M. Monkman, *Peer review in scholarly journals : perspective of the scholarly community - an international study*, Londres, Publishing Research Consortium, 2008.
- [15] T. Jefferson, E. Wager et F. Davidoff, « Measuring the Quality of Editorial Peer Review », *JAMA*, vol. 287, n°21, 2002,

p. 2786-2790.

- [16] G. D. L. Travis et H. M. Collins, « New Light on Old Boys : Cognitive and Institutional Particularism in the Peer Review System », *Science, Technology, & Human Values*, vol. 16, n° 3, 1991, p. 322-341.
- [17] Les prénoms des chimistes interrogés ont été modifiés.
- [18] B. Milard, « La soumission d'un manuscrit à une revue : quelle place dans l'activité scientifique des chercheurs ? » *Schedae (Presses Universitaires de Caen)*, n° 1, 2008, p. 1-12.
- [19] R. Park, « La ville. Propositions de recherche sur le comportement humain en milieu urbain » [1925] in *L'école de Chicago. Naissance de l'écologie urbaine*, Y. Grafmeyer et I. Joseph, Paris, Aubier, 1984, p. 83-130, p. 121.
- [20] G. Mallard, M. Lamont et J. Guetzkow, « Fairness as Appropriateness : Negotiating Epistemological Differences in Peer Review », *Science Technology Human Values*, 34, 2009, p. 573-606.
- [21] R. Park, *op.cit.*, p. 121.
- [22] R.K. Merton, 1968. « The Matthew Effect in Science : The reward and communication systems of science are considered », *Science*, 159, n° 3810, p. 56-63.
- [23] M. W. Rossiter, « The Matthew Matilda Effect in Science », *Social Studies of Science*, vol. 23, n° 2, 1993, p. 325-341.
- [24] D. Crane, « The gatekeepers of sciences : some factors affecting the selection of articles of scientific journals », *American sociologist*, vol. 2, n° 1, 1967, p. 195-201.
- [25] M. Ware et M. Monkman, *op. cit.*
- [26] D.E. Chubin et E. J. Hackett, *Peerless Science : Peer Review and US Science Policy*, State University of New York Press, New York, 1990.
- [27] S. Jasanoff, « Peer Review in the Regulatory Process », *Science Technology Human Values*, vol. 10, n° 3, 1985, p. 20-32.
- [28] Discours N. Sarkozy, « Discours à l'occasion du lancement de la réflexion pour une stratégie nationale de recherche et d'innovation », <http://www.elysee.fr/president/les-&hellip;> ; (page consultée le 22 octobre 2010).
- [29] Y. Gingras, « Du mauvais usage de faux indicateurs », *Revue d'Histoire Moderne et Contemporaine*, vol 5, n° 55-4bis, 2008, p. 67-79.
- [30] M. Grossetti et B. Milard, « Note sur les indicateurs bibliométriques en SHS », document de travail, Toulouse, 2009.
- [31] N. C. Mullins, « The development of a scientific specialty : the phage group and the origin of molecular biology », *Minerva*, vol. 10, n° 1, 1972, p. 51-82.
- [32] K. Kousha et M. Thelwall, « Google Scholar citations and Google Web/URL citations : A multi-discipline exploratory analysis », *Journal of the American Society for Information Science & Technology*, n° 58, 2007, p. 1055-1065