



HAL
open science

Redouter l'évaluation?

Denis Jerome

► **To cite this version:**

Denis Jerome. Redouter l'évaluation?. Jahrbuch für Computerphilologie, 2008, n°10, pp.18. hal-00363292

HAL Id: hal-00363292

<https://hal.science/hal-00363292>

Submitted on 21 Feb 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Redouter l'évaluation ?

Denis Jérôme* (jerome@lps.u-psud.fr)

Laboratoire de Physique des Solides, Université Paris-Sud, 91405 Orsay

De toute part émergent des propositions pour évaluer les scientifiques à partir des données bibliométriques de leurs publications. Ces nouvelles méthodes peuvent être utilisées, dans la mesure où elles sont directement liées à la reconnaissance des auteurs par leurs citations, au travers de leur indice de Hirsch ou de Egghe et non à la popularité du journal (facteur d'impact). Un petit exercice basé sur des scientifiques reconnus et de différentes communautés montre comment l'indicateur de Egghe, en particulier, peut être un complément utile à l'indispensable évaluation par les pairs.

Dès lors que le scientifique demande l'aide de l'État pour financer postes et projets, ne se trouve-t-il pas dans l'obligation de se soumettre à l'évaluation des résultats de ses recherches ? La société souhaite désormais une échelle de valeurs lui permettant d'évaluer et d'établir une hiérarchie, tant pour les personnels que pour les laboratoires ou les performances des différents pays. En témoigne le succès médiatique remporté par les classements des universités élaborés par différents organismes [1], mais qui ne sont cependant pas pertinents dans notre pays, compte tenu de la structure très particulière de notre recherche.

L'activité des scientifiques des pays les plus industrialisés se matérialise chaque année par environ un million de publications. Le rôle de ces publications est multiple [2] : informer le grand public des avancées scientifiques, contribuer aux progrès ultérieurs du développement de la recherche, et enfin constituer un des supports de l'évaluation de la recherche avec toutes ses composantes. Ce dernier rôle, autrefois marginal, est devenu l'une des préoccupations majeures des chercheurs, et ceci, à cause des effets conjugués de la professionnalisation de la recherche et de l'intrusion de la loi du marché dans la diffusion scientifique [3].

L'évaluation de la recherche est certainement un exercice des plus délicats, dans la mesure où elle mêle des considérations à la fois objectives et subjectives. Elle est pratiquée par un comité qui fonctionne généralement suivant la formule de l'évaluation par les pairs. Toutefois, évaluer relève souvent de la quadrature du cercle, étant donné que le comité doit être compétent dans le domaine tout en restant désintéressé par le résultat de ses conclusions. Au niveau des personnels, il apparaît normal d'utiliser le travail du chercheur, matérialisé par ses publications, comme support de cette évaluation. Il convient néanmoins de tenir compte du fait que la contribution d'un chercheur au progrès scientifique peut prendre des formes très diverses, qu'il est parfois délicat d'appréhender (participation à l'enseignement et à l'encadrement de thèses, rédaction d'ouvrages, vulgarisation, administration, rayonnement international, prix et récompenses internationales, etc.). De plus, le poids des publications dans le jugement est variable selon qu'il s'agit d'un jeune postulant à un emploi au CNRS ou à l'enseignement

supérieur, dont le dossier n'est riche que d'un petit nombre de publications récentes, ou du chercheur confirmé dont le nombre de travaux peut dépasser aisément la centaine, ou encore du patron qui cosigne tous les travaux provenant de son laboratoire.

Utiliser les publications comme complément à l'évaluation des chercheurs revient à prendre en compte le jugement par les pairs à deux reprises : d'abord au niveau de la revue, la décision de publier dépend de l'éditeur après avis des experts ; ensuite, au niveau de l'intérêt suscité par l'article qui prend la forme de citations dans les travaux ultérieurs. À l'indispensable évaluation par les pairs vient s'ajouter désormais une évaluation complémentaire basée sur la bibliométrie. En effet, le fantastique développement du *web* a bouleversé les pratiques de la publication scientifique [4]. Il permet notamment l'accès à des bases de données qui fournissent les informations sur les citations relatives aux articles publiés dans les revues.

Le facteur d'impact des revues scientifiques

Une première utilisation des publications dans l'évaluation des chercheurs a été faite sur un critère de popularité des revues dans lesquelles un chercheur a publié ses travaux, partant du principe que plus un journal est cité, plus le chercheur qui parvient à y être publié mérite d'être reconnu. Le prestige d'une revue est mesuré par son « facteur d'impact », qui a été et demeure encore très utilisé pour l'évaluation des personnels ou des organismes⁽¹⁾. Il est normal dans ce contexte, que les chercheurs tentent d'être publiés par des revues à facteur d'impact élevé afin de s'approprier un peu de leur prestige.

Le facteur d'impact des revues doit être considéré avec beaucoup de précaution, car il peut induire en erreur. En effet, des magazines pluridisciplinaires comme *Nature* ou *Science* possèdent un facteur d'impact de l'ordre de 30, alors qu'un facteur de 6 ou 7 pour un journal spécialisé est déjà une performance remarquable. Ce facteur d'impact équivaut sommairement à un indice de citation moyen annuel des articles de la revue. Cependant, il faut garder en mémoire qu'il existe un pic très marqué pour un nombre restreint d'articles à taux de citations élevés⁽²⁾. En conséquence, ce taux moyen de

* L'auteur, directeur de recherches au CNRS, a été au cours des 30 dernières années fondateur et éditeur en chef de plusieurs périodiques européens en physique, et fait partie de comités de réflexion sur les revues scientifiques à la Société Française de Physique et sur l'évaluation à l'Académie des sciences.

citations peut, suivant les domaines, s'éloigner fortement du taux moyen d'un domaine en particulier. D'ailleurs, si les auteurs pensent optimiser leur visibilité en recherchant à tout prix la publication dans des magazines multidisciplinaires à facteur d'impact élevé, il font probablement fausse route. Comme le montre une étude récente [5], les articles du domaine de la physique, par exemple, sont plus visibles et cités lorsqu'ils sont publiés dans des revues spécialisées comme *Physical Review (Letters)* plutôt que dans les magazines tels que *Nature* ou *Science*. Et même Karl Ziemelis, éditeur en chef à *Nature*, commente l'utilisation du facteur d'impact dans les termes suivants [6] :

It never ceases to amaze me the reverence with which journal impact factors are held or, more worryingly, the widespread misuse of these quantities as metrics for assessing the research output of individual scientists. At most, an impact factor is a crude measure of the "success" of a journal in attracting highly cited papers; it is a time-dependent average of citations over all of that journal's content, and so says essentially nothing about the strengths or weaknesses of individual research papers.

Autre effet pervers de l'utilisation du facteur d'impact des revues dans l'évaluation : des travaux réellement originaux rencontrent souvent plus de difficultés à être publiés dans les magazines pluridisciplinaires que ceux qui se situent dans des courants de modes. Et comme l'écrivait Pierre Joliot [7] : « un projet d'article était aisément accepté s'il s'agissait d'un travail sérieux mais dans la norme ». L'évaluateur doit donc rester conscient que son rôle n'est pas de juger du prestige d'un journal dans lequel un travail est publié, mais bien de la qualité d'un article publié dans un journal donné. Si le facteur d'impact doit être banni du processus d'évaluation des personnels, comment alors juger le plus objectivement possible à la fois de la productivité d'un auteur et de la qualité de son travail ?

Les indicateurs de citations

Les possibilités de bibliographie inversée offertes par les bases de données actuelles permettent désormais le calcul de divers indicateurs liés à l'activité scientifique des personnes⁽³⁾. Il en est ainsi pour un indicateur suggéré en 2005 par J. Hirsch, qui a proposé une mesure de la quantité et de la qualité scientifique par l'indice h [8]. Cet indicateur est devenu rapidement très populaire et tend même à devenir un outil classique de la bibliométrie. L'indicateur h est un nombre, associé à un chercheur et calculé en considérant l'impact de l'ensemble de ses publications (mais il peut également être étendu à d'autres ensembles comme un groupe de chercheurs, un laboratoire, un domaine, un pays, des journaux, etc.). Il est

Nombre de citations (C_i)

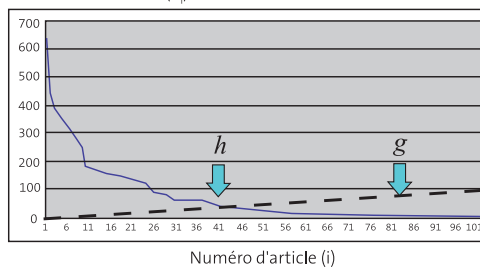


Figure 1. Les indicateurs h et g sont illustrés par l'exemple de la figure. La courbe bleue représente le nombre (décroissant) C_i de citations en fonction du numéro (i) d'article. La droite en tirets correspond à $C_i = i$. Les deux flèches indiquent les valeurs des indicateurs de Hirsch ($h=41$) et de Egghe ($g=83$).

déterminé en classant les articles d'un auteur en fonction du nombre décroissant, C_i , de leurs citations, de sorte que le papier de rang $i=h$ ait reçu au minimum h citations (fig. 1). Cet indicateur présente l'intérêt de mettre en valeur les articles qui sont les plus cités. Il tient compte de tous les papiers très cités, sans donner un avantage aux poids lourds que représentent parfois des articles de revue parus dans des publications prestigieuses comme *Review of Modern Physics* en physique ou *Chemical Reviews* en chimie. Il existe une bonne corrélation entre h et la production globale d'un auteur ainsi qu'avec le jugement par les pairs [9]. Cet indicateur dépend évidemment de l'âge du chercheur concerné et permet de tenir compte de différentes séniorités en calculant un facteur h normalisé par le nombre d'années d'activité professionnelle.

Ce qui a rapidement contribué à la popularité de cet indicateur, c'est sa grande facilité de calcul à partir des bases de données bibliographiques. Toutefois, comme pour tous les indicateurs bibliométriques, le problème principal laissé à la charge de l'évaluateur réside dans la pertinence du fichier bibliographique de départ, avec la suppression des homonymies, le problème des noms composés et le bon choix des domaines scientifiques.

Cependant, l'indicateur h souffre de nombreuses insuffisances. Il n'est pas adapté, par exemple, au chercheur qui n'aurait publié qu'un nombre restreint d'articles tout en ayant attiré l'attention de nombreux autres auteurs, puisqu'il peut ne pas être défini dans ce cas. De plus, l'indicateur h ne permet pas de discriminer entre deux profils de chercheurs de même indicateur : l'un avec un total modeste de citations jusqu'à l'article de rang h et l'autre présentant un total élevé de citations pour les h premiers articles. En outre, l'indicateur n'évolue pas lorsque ce sont les h premiers articles qui voient leurs citations augmenter au cours du temps. C'est souvent le cas, puisque les articles les plus cités jouissent *de facto* d'une plus grande visibilité, une situation qui accroît encore plus

Références

1. Voir, par exemple, le classement de Leiden : <http://www.cwts.nl/cwts/LeidenRankingWebSite.html> ou le classement de Shanghai : http://ed.sjtu.edu.cn/rank/2007/ARWU2007_TopEuro.htm
2. P.A. Lawrence, *Ethics in Science and Environmental Politics*, vol 8, open access preprint 2008, doi: 10.3354/esepp00079 <http://www.int-res.com/journals/esepp/esepp-home/>
3. D. Jérôme, « L'autorité des grandes revues scientifiques », dans les actes du colloque sur *L'autorité* publiés sous la direction d'Antoine Compagnon, Éditions Odile Jacob, à paraître (2008).
4. Voir sur ce point le remarquable numéro de *Physics World*, janvier 2007.
5. C.W. Miller, "Superiority of the h-index over the Impact Factor for Physics", <http://arxiv.org/abs/physics/0608183v1>
6. Communication personnelle de K. Ziemelis.
7. P. Joliot, *La recherche passionnément*, Éditions Odile Jacob, Paris, 2001.
8. J. Hirsch, "An index to quantify an individual's research output", *Proc. Nat. Acad. Sci. USA*, **102** (2005) 16569, ou <http://arxiv.org/abs/physics/0508025>
9. A.F.J. van Raan, "Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups", <http://arxiv.org/abs/physics/0511206>
10. R.K. Merton, "The Matthew effect in Science", *ISIS*, **79** (1988) 606-623.
11. L. Egghe, "An improvement of the h-index", *ISSI Newsletter*, **2**(1), 8-9 et "Theory and practice of the g-index", *Scientometrics*, **69** (2006) 131.
12. <http://pasquier.claude.free.fr/publications/publisdata.php>

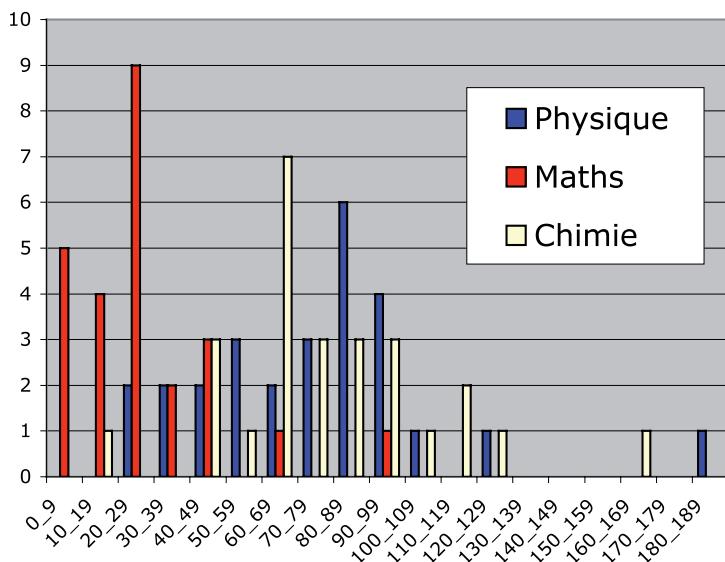


Figure 2. Distribution des mathématiciens, physiciens et chimistes de l'Académie des sciences en fonction de leur indicateur g . Les résultats de cette étude ont été obtenus avec la fenêtre de dates 1975-2007, sauf pour quelques membres plus anciens pour lesquels la fenêtre débute en 1945.

(1) Le facteur d'impact d'une revue X pour l'année N est défini par la base de données ISI-Thomson (<http://scientific.thomson.com>), qui indexe plus de 6000 périodiques en science. Il est égal au rapport A/B entre le nombre total (A) de citations à la revue X dans l'ensemble des revues enregistrées dans la base ISI durant l'année N des articles parus en N-1 et N-2 dans la revue X, et le nombre total (B) des articles publiés par la revue X pendant ces mêmes années N-1 et N-2. Notons que la période d'examen de deux ans est totalement arbitraire et ne tient pas compte des différences de cultures entre domaines.

(2) P.O. Seglen, *British Medical Journal*, 314 (1997) 498, fait remarquer que pour un journal classique, environ 15% de ses articles contribuent à la moitié des citations engrangées par le journal.

(3) Le Web of Science, payant, est géré par ISI-Thomson, <http://scientific.thomson.com/>. Scopus, <http://www.scopus.com/>, est un serveur payant de l'éditeur Elsevier. Google Scholar, <http://scholar.google.fr/> et ADS, <http://cdsads.u-strasbg.fr/> sont des serveurs d'accès gratuit.

(4) La question de la liste de références dans un article représente un problème sérieux, auquel trop peu d'attention est prêtée en général par les rapporteurs et les rédacteurs en chef. Il y a souvent trop de références de complaisance qui n'ont pas de raison d'être, alors que d'autres, concernant des travaux de base, ne figurent pas.

(5) Le cas des mathématiques est une excellente illustration de la spécificité de certains domaines (voir J.M. Schlenker, *Gazette de la SMF*, p. 73, janvier 2008). La durée de vie d'un article en mathématiques pouvant largement excéder 3 ans, l'évaluation des individus ne peut se faire de façon fiable sur la base des citations faites à leurs travaux récents.

la possibilité de les citer, en quelque sorte, une manifestation du phénomène de l'avantage cumulé symbolisé par le principe de Matthieu [10].

Certains des défauts de l'indicateur de Hirsch ont toutefois été corrigés par L. Egghe, qui a proposé un indicateur g privilégiant la qualité des travaux scientifiques par rapport à leur quantité [11]. La valeur de g est telle que le total des citations reçues par les g premiers articles soit supérieur ou égal au carré de g , c'est-à-dire si les deux conditions suivantes :

$$\sum_{i=1}^{i=g} C_i \geq g^2 \quad \text{et} \quad \sum_{i=1}^{i=g+1} C_i < (g+1)^2 \text{ sont satisfaites.}$$

L'indicateur g reflète bien mieux que ne le fait h l'impact des travaux même peu nombreux mais magistraux et de grand intérêt pour le développement scientifique⁽⁴⁾. Cet indicateur n'est jusqu'à présent pas calculé par les bases de données et il demande donc un travail à partir de chaque fichier bibliométrique. Ce travail est cependant rendu moins fastidieux par l'utilisation d'un logiciel mis au point par C. Pasquier au laboratoire LPS d'Orsay et en accès libre sur le web [12].

Du bon usage des outils d'évaluation bibliométriques

Ayant donc à notre disposition quelques outils bibliométriques, la question se pose de savoir si ces outils doivent être utilisés et dans quelles conditions. Il est tout d'abord évident que ces indicateurs perdent toute signification si l'exercice n'est pas effectué sur un ensemble homogène de chercheurs comparables, tant par leur domaine scientifique que par leur classe d'âge.

À titre d'exercice pratique, un exemple, basé sur l'indicateur g de tous les membres des sections de mathématiques, de physique et de chimie à l'Académie des sciences, fournit

l'histogramme de la figure 2. Cette distribution indique un g moyen croissant en allant des mathématiques vers la physique, situation conforme aux coutumes des différents domaines. L'exercice pourrait être mené sur des ensembles plus larges tels que les chercheurs du CNRS ou d'autres organismes. Une autre possibilité ouverte par ces indicateurs est le suivi de l'évolution des chercheurs en réactualisant les données d'une année sur l'autre.

Ce petit exercice, pratiqué sur des ensembles homogènes de chercheurs, valide dans une certaine mesure la fiabilité de l'outil.

Il existe cependant des limites à l'emploi des indicateurs bibliométriques. Alors que trois des quatre derniers Nobel de physique (dont le g moyen est de 105) appartiennent au premier tiers des membres dans un classement par g décroissant, les Nobel de chimie se situent aux deux extrémités de la distribution. Les mathématiques, quant à elles, font ressortir un g moyen qui n'est que de 38 pour les médailles Fields. Ces nombres illustrent les différences de tailles des communautés, ainsi que leurs différences d'habitudes en matière de politique de publications⁽⁵⁾.

Rappelons que le facteur d'impact des revues doit être proscrit de l'arsenal de l'évaluateur. En outre, en aucun cas les indicateurs de citations ne peuvent se substituer à un examen approfondi de quelques publications proposées à l'évaluation et sélectionnées par les candidats eux-mêmes. De plus, il faut se garder de prendre pour argent comptant la valeur de l'indicateur. Leur utilité doit se limiter à apporter un peu d'objectivité dans l'exercice délicat de l'évaluation, en s'adossant au jugement par les pairs et ne doit jamais dispenser de la consultation du fichier des publications associées aux citations qui ont servi à déterminer l'indicateur. La constatation d'un indicateur situé dans la tranche basse de la distribution doit éveiller l'attention et engager à un examen encore plus approfondi du dossier des publications. Cette situation peut fort bien arriver pour un chercheur qui aurait publié l'essentiel de ses résultats dans des livres ou des ouvrages non répertoriés par ISI. Dans le cas opposé d'un indice trouvé particulièrement élevé, il y aura aussi lieu d'examiner de près la contribution du chercheur à des travaux abondamment cités. Moyennant la prise en compte de toutes ces précautions, la bibliométrie effectuée en toute transparence peut devenir une aide à l'évaluation. ■

Remerciements

Je remercie Jacques Friedel pour de nombreuses discussions et ses commentaires pertinents, ainsi que mes confrères du groupe « Évaluation » de l'Académie des sciences, et tout particulièrement Marc Yor pour ses remarques sur le domaine des mathématiques.