

**Remèdes potentiels à l'inadéquation des méthodes
d'évaluation quantitative appliquées aux sciences
humaines et sociales**

Manuel Durand-Barthez

► **To cite this version:**

Manuel Durand-Barthez. Remèdes potentiels à l'inadéquation des méthodes d'évaluation quantitative appliquées aux sciences humaines et sociales. Bulletin de psychologie, Groupe d'étude de psychologie, 2015, 68(5) (539), pp.415-417. sic_01426153

HAL Id: sic_01426153

https://archivesic.ccsd.cnrs.fr/sic_01426153

Submitted on 4 Jan 2017

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.



REMÈDES POTENTIELS À L'INADÉQUATION DES MÉTHODES D'ÉVALUATION QUANTITATIVE APPLIQUÉES AUX SCIENCES HUMAINES ET SOCIALES

Manuel DURAND-BARTHEZ

Conservateur de bibliothèque, co-responsable de l'Unité régionale de formation à l'information scientifique et technique (URFIST) de Paris

À travers le programme P150 de la LOLF (Loi organique des lois de finance) visant le financement de la Recherche universitaire, l'Administration centrale préconise l'utilisation d'indicateurs de la production scientifique à partir d'une méthode bibliométrique. Celle-ci est plaquée sur un modèle anglo-saxon touchant essentiellement les Sciences de la Vie et de la Matière. Outre les biais de ces méthodes ici évoqués, il apparaît que leur inadéquation aux Sciences humaines et sociales suscite un profond malaise dans ce secteur en général et singulièrement dans celui de la Psychologie. Est enfin abordé le problème de la langue anglaise dont l'usage quasi obligatoire dans les publications rendrait l'élaboration de celles-ci peu compatibles avec les conditions d'exercice des chercheurs praticiens.

POTENTIAL REMEDIATION TO QUANTITATIVE EVALUATION METHODS APPLIED TO HUMAN AND SOCIAL SCIENCES

Through the P150 Program of the French Organic Financial Law, concerning the funding of the University Research, the central government recommends the use of indicators measuring the scientific production, based upon bibliometrics. It is akin to an Anglo-saxon model affecting essentially Life Sciences and Physics. Besides the biases of this method, which are here described, its inadequacy with Human and Social Sciences induces a deep malaise in these fields in general, and with Psychology in particular. In conclusion, the subject of the English language is broached, the use of which is almost mandatory in publications. In the French field of Psychology, this is rather inconsistent, especially with the conditions of practice of the clinicians.

TEXTE

Voilà déjà une dizaine d'années, dès 2003, les premiers articles publiés dans *Nature* au sujet de l'évaluation ont paru, signés par des chercheurs en biologie moléculaire et cellulaire, de l'*University College of London* et de *Cambridge*, sur la « tyrannie du facteur d'impact », sur tous les biais et toute la perversité du mode d'évaluation actuelle. Le débat a été repris plus récemment par Yves Langevin, responsable de l'Institut d'astrophysique du CNRS, et par Denis Jérôme, chercheur au Laboratoire de physique des solides d'Orsay et co-auteur du rapport de l'Académie des sciences du 17 janvier 2011 sur le mauvais usage de la bibliométrie.

Conférence des publications de psychologie en langue française. Journée du 28 septembre 2013.

Rôles et usages des publications scientifiques en psychologie, quels enjeux pour les chercheurs et les praticiens ?

Des scientifiques de haut rang, directement concernés par ces questions, sont les premiers à dire et à mettre sur la table le fait que tout cela est profondément pervers et même qu'« on le dit, mais qu'on ne fait rien ». C'est dire qu'il y a une *omerta* sur cette question, suscitée par un certain nombre de pressions. Il faut rappeler qu'en France, la loi organique relative aux lois de finances (LOLF) se déploie, pour la recherche, sur un certain nombre de programmes. La recherche universitaire, en particulier, est concernée par le « Programme 150 - Formations supérieures et recherche universitaire » (P150) qui assortit la production des acteurs de la recherche de deux formules : la « justification au premier euro » et, « du point de vue du contribuable ». Trois tableaux se réfèrent aux pourcentages de citations exigibles de la part des acteurs de la recherche, avec des cibles qui s'étendent année après année.

Le P150 a été mis au point en 2008 et renouvelé en 2012 ; il est doté d'un certain nombre de projections. Le P150 est fondé sur une base de données à caractère plus ou moins monopolistique : le *Web of Science*. C'est écrit en toutes lettres : c'est la référence. Le *Web of Science* sert à l'Observatoire des sciences et techniques (OST) qui est un Groupement d'intérêt public (GIP), chargé par l'État de mettre au point des indicateurs de l'évaluation de la recherche. L'OST fonctionne à partir du *Web of Science*. Celui-ci a mis au point deux types d'indicateurs importants : le premier est le facteur d'impact, que vous connaissez bien, et qui consiste à évaluer un contenu à partir de son contenant, c'est-à-dire à partir de la revue. Or, il ne faut pas se leurrer, il n'existe pas que le facteur d'impact des revues, il existe aussi le facteur d'impact agrégé de ce qu'on appelle les *Subject Categories*, réparties sur le *Science Citation Index-Expanded* (SCI-E) pour le domaine Sciences-Technologie-Médecine, avec un ensemble de 173 disciplines bien définies, et le *Social Sciences Citation Index* (SSCI), avec 56 *Subject Categories*. Ces *Subject Categories* contiennent un certain nombre de titres de revues et l'on va calculer le facteur d'impact agrégé de chacune de ces catégories. La cotation des catégories varie de la même manière que la cotation des revues. De ce fait, certaines institutions, au moment des bilans annuels, vont dire : « Vous nous mettez dans cette catégorie, mais nous demandons à migrer sur telle autre catégorie dont la technicité est relativement connexe, parce que son facteur d'impact agrégé est préférable ». Cela signifie que, dans la mesure où les *Subject Categories* du *Web of Science* sont reprises intégralement par l'OST, il est clair qu'une agence de moyens ou un bailleur de fonds institutionnel peut, au reçu d'un certain nombre de dossiers, dire que l'un d'eux relève de telle ou telle catégorie dont la cote, actuellement, semble être en hausse et, que par conséquent, il paraît plus opportun de favoriser ce genre de proposition.

Le facteur H, de la même manière, a été mis au point pour l'évaluation des auteurs – un facteur H de 25, signifie qu'au moins 25 des publications de cet auteur incluses dans le *Web of Science* sur une période donnée, sont citées au moins 25 fois. Les biais du facteur H ont été dénoncés radicalement par toute la communauté scientifique depuis des années.

On nous dit que le facteur d'impact ne vaut rien, que le facteur H ne vaut rien, mais, en attendant, on continue. Le facteur H ne concerne pas que les individus. Le facteur H concerne également les revues. Il existe un facteur H des revues et, plus encore, un facteur H des pays. Tout ceci a tendance à favoriser une autonomie stratégique des revues et partant, une manière de mesurer que l'on qualifie communément de *Quick and Dirty*. *Quick* c'est une évaluation quantitative rapide. Arnold Migus, alors qu'il dirigeait le CNRS, disait : « Donnez-moi la liste des facteurs H, j'en ferais bien mon affaire pour les promotions. » Le facteur H et, par conséquent, le facteur d'impact qui sont ici concernés, sont beaucoup plus adaptés aux Sciences de la Vie et de la Matière, et toute la communauté scientifique le reconnaît. En effet, dans ces deux domaines existent une culture de publication et une culture de citation qui favorisent ce genre de chose. L'évaluateur administratif, à son plus haut niveau, par exemple

lorsqu'il s'agit du Grand emprunt, s'appuie sur un constat simple: « La preuve du travail fourni par une équipe de recherche est ce qui est visible », et ce qui est visible, c'est ce qui est rendu public, donc publié, d'où le *Publish or Perish*. Dans la mesure où les évaluateurs administratifs n'ont ni le temps ni la compétence pour aller en profondeur dans ces publications, le principe de la citation, c'est-à-dire de la capitalisation des citations, va permettre d'effectuer ce travail rapide de l'estimation, à la fois, par le facteur d'impact et le facteur H.

D'autres systèmes existent. Mon collègue¹ citait le cas du responsable de laboratoire qui figure en dernière position dans la liste des signataires d'un article. Ceci correspond à certaines pratiques de publication dans le milieu de la recherche biomédicale, comme le Système d'interrogation, de gestion et d'analyse des publications biomédicales (SIGAPS) qui a été mis en œuvre dans les établissements de recherche du secteur Médecine. Au moment de l'évaluation annuelle, un logiciel va prendre en compte l'auteur numéro 1, considéré souvent comme étant le thésard, le 2, souvent son encadrant, et le dernier auteur, qui est le « patron ». Le signataire classé N+3, ou avant-dernier, sera classé avec une moindre valeur. À l'Inserm, par exemple, il y a parfois des négociations entre le patron et l'avant-dernier auteur sur le thème: « peux-tu céder ta place cette année ? ». Ces pratiques sont assez courantes. En physique des particules, un même type de calcul est effectué, à partir d'un autre ordre. Dans les universités et les CHU, des logiciels composés de trois volets sont utilisés. Le premier volet traite la gestion budgétaire et comptable de l'unité de recherche ; le deuxième volet concerne les dossiers de carrière des personnels, les enseignants-chercheurs, doctorants, personnels ITRF ou BIATSS ; le troisième volet traite les publications. Entre ces trois volets, un principe de vases communicants est à l'œuvre.

Pour la recherche en sciences humaines, de quelle documentation, traduisant l'esprit de l'administration centrale dans ce domaine, dispose-t-on ? On note par exemple la réflexion sur l'évaluation préalable au Grand emprunt, commise par le Conseil pour le développement des humanités et sciences sociales (CDHSS), et le second rapport d'étape remis à M^{me} Valérie Péresse le 10 mars 2010. Dans ce rapport, on reconnaît les biais de la bibliométrie ; on reconnaît que la bibliométrie actuelle est essentiellement adaptée aux Sciences de la Vie et de la Matière, et absolument pas aux Sciences humaines et sociales. On le reconnaît, mais, aujourd'hui, nul ne sait trop bien quoi proposer. On peut y lire ce constat des biais sur plusieurs pages. Il est mentionné que l'AERES s'efforce d'appliquer des mesures adaptées – ils font des efforts, mais c'est difficile – et que le CDHSS se prononce pour une évaluation multidimensionnelle. Il n'est pas rare d'entendre aujourd'hui: « un chercheur en sciences humaines et sociales a plusieurs cordes à son arc, et il peut aussi bien rédiger des articles qu'un chapitre de livre, des tutoriels, des cours en ligne, des catalogues d'expositions, organiser des expositions, réaliser des films, des vidéos, faire des traductions, avoir une compétence polyglossique, et tout cela peut être consigné sur les fiches d'identification du chercheur ». Toutes ces modalités de travail, autres que l'article paru dans un périodique de rang A, qui est évidemment privilégié par l'Administration centrale, sont susceptibles d'être prises en compte. On ne sait pas si ceci va effectivement peser face aux critères fondamentaux qui sont ceux du P150 et qui régissent les évaluations brutes, par exemple de l'OST, et tout ce qui a pu présider à l'organisation du Grand emprunt. L'INSHS-CNRS, qui s'est prononcé de façon extrêmement forte, également en juin 2011, sur l'évaluation, contre la bibliométrie telle qu'elle est conçue, a mis au point une base de données dénommée *Journal Base*, qui collecte

¹ voir ci-dessus, l'intervention de Felice Carugati (NDLR).

dix milles titres de périodiques en sciences humaines et sociales, et qui rassemble à cette fin les fichiers du *Web of Science*, bien sûr, les fichiers de SCOPUS qui est le produit du « concurrent n°1 » de Thomson-Reuters (concepteur du *Web of Science*) : c'est-à-dire Elsevier, et les deux fichiers de l'ERIH (*European Reference Index for the Humanities*) et de l'AERES, et là, c'est une classification par base de données. Qui va empêcher un bailleur de fonds institutionnel de dire : « Cette revue A est dans le JCR (*Journal Citation Reports*) qui est le fichier situé dans l'abaque du *Web of Science*, donc je vais préférer financer de ce côté-là » ? Le problème de l'orientation des bases est aussi difficilement appréciable. D'autres critères entrent en jeu : l'« Altmetrics » consiste à faire des évaluations à partir de la répercussion des travaux scientifiques sur les réseaux sociaux. Dans quelle mesure tel chercheur, telle équipe, tel article se manifestent-ils par les biais de Tweeter, de Face Book ou de blogs divers ; mais aussi comment les carnets de recherches de *Revues.org* coproduits par l'EHESS vont-ils pouvoir être comptabilisés ? Ce sont des choses qui commencent à émerger, mais il est clair que les institutions, en particulier les administrations centrales, y sont très réticentes.

Il reste la question de la langue, qui a été abondamment évoquée tout à l'heure. Quelle est, aujourd'hui, la position du département Stratégie et innovation - secteur SHS - de la DGRI, c'est-à-dire de l'administration centrale ? Cette position est précisée par l'Alliance ATHENA, réunissant un certain nombre de regroupements ou fédérations de centres de recherches concernant les sciences humaines et sociales, et par l'Alliance nationale pour les sciences de la vie et de la santé (AVIESAN) qui est très puissante, et qui, évidemment, va dans le sens de la bibliométrie classique. L'Alliance ATHENA a commis, en septembre 2012, un rapport sur l'internationalisation de la recherche en Sciences humaines et sociales, dans lequel l'usage quasi exclusif de l'anglais est préconisé. Sachant qu'aujourd'hui, ce sont des sujets qui sont en cours de discussion à la DGRI, il s'agit de trouver des niches qui permettent de récupérer de l'argent pour financer les équipes de rédacteurs, afin d'inciter les chercheurs en sciences humaines et sociales à faire, eux-mêmes, une première ébauche en anglais plus ou moins adroite, qui serait reprise par des réviseurs professionnels, lesquels coûteraient moins cher que les traducteurs à part entière. Dans la politique de l'Administration centrale, aujourd'hui, c'est un encouragement vif et très fort à l'internationalisation. Je mentionnerai également des rapports qui sont beaucoup plus conséquents dans leur analyse que le rapport français que je citais tout à l'heure, par exemple celui de l'Académie royale néerlandaise des arts et des sciences, sur l'évaluation de la recherche en sciences humaines ; un rapport de 68 pages paru en mai 2011 « *Quality indicators for research in the humanities* », six ans après un premier rapport « *Judging research on its merits* » qui épluchait tous les biais de l'évaluation quantitative en sciences humaines. Leurs auteurs mettent en avant l'impact sociétal, la *societal priority*, qui, aujourd'hui, semble très important dans les politiques prospectives en matière d'évaluation des sciences sociales : faire une évaluation en fonction des impacts de la recherche sur la société, c'est-à-dire ce qui est tangible en vue d'améliorer le bien-être de l'humanité. Ce qui pourrait, dans le contexte de la psychologie, correspondre au concept de recherche translationnelle, c'est-à-dire voir comment la recherche fondamentale dans le domaine médical (mais c'est la même chose en psychologie) peut transposer ses effets futurs sur la société. Un certain nombre de critères peuvent éventuellement être élaborés pour mettre en évidence ce type d'indication, tant en matière de vulgarisation qu'en regard de la mise en place de structures d'aide sociale, d'insertion sociale, etc. Ce critère de l'impact sociétal serait l'une des pistes sur lesquelles l'Administration centrale pourrait se prononcer, sachant qu'au sein des sphères d'influences, il se dit que le secteur SHS représente « la cinquième roue du carrosse dans toutes les instances qui décident de la politique de la recherche en France ». À l'occasion du Grand emprunt, ou autres événements de ce genre, il est clair que les

représentants, même au plus haut niveau de l'Administration centrale en sciences humaines et sociales, sont significativement marginalisés.

Pour terminer, l'un des produits dérivés du *Web of Science* est l'*ISI Highly Cited* qui est censé recenser les 250 meilleurs chercheurs au monde les plus cités, en fonction de 21 disciplines, lesquelles sont reprises par l'OST. Sur ces 21 disciplines, seulement 3 recouvrent les Sciences sociales : l'une s'appelle *Social Sciences General*, une autre : *Economics Business* et une troisième : *Psychiatry /Psychology*. Ce qui veut dire que la psychologie est essentiellement associée à des considérations médicales, derrière lesquelles il y a, bien évidemment, les gros laboratoires pharmaceutiques intéressés par les neuroleptiques, les psychotropes etc.