



Productivité et qualité scientifique : avec quelles publications compter ?

par Olivier Rey

Depuis plusieurs mois, une partie du débat sur l'université porte sur les questions d'évaluation des activités scientifiques et plus particulièrement sur l'appréciation du travail des chercheurs en fonction de leurs publications. Classements de revues, bibliométrie, indices de citation, palmarès, indicateurs de qualité : la polémique, à la fois nationale et internationale, porte sur des notions souvent mal connues dans le monde des sciences humaines et sociales.

Le cœur de la controverse réside surtout dans l'idée que l'on puisse utiliser des mesures quantitatives opérées sur des publications pour apprécier la productivité scientifique des chercheurs et enseignants-chercheurs. Si la communauté scientifique est plutôt attachée aux dispositifs d'évaluation par les pairs (*peer review*), l'évaluation par les chiffres fascine en effet toujours les décideurs et les profanes soucieux de disposer de références simples à comprendre et à manipuler.

Par ce dossier, nous souhaitons resituer le sujet dans le cadre de l'évaluation des universités mais aussi fixer un certain nombre de repères précis concernant la recherche en éducation, pour contribuer à éclaircir les enjeux du débat. On y trouve donc une synthèse des sources indispensables à la compréhension des arguments échangés, complétée par des références bibliographiques récentes permettant d'aller plus loin.

[Les universités dans les nouvelles vagues de l'évaluation](#) | [La querelle des classements de revues : l'exemple de la recherche en éducation](#) | [Évaluation par les pairs vs évaluation par les chiffres](#) | [Comment évaluer la qualité en SHS ?](#) | [Conclusion](#) | [Bibliographie](#).

Avertissements au lecteur :

- *la plupart des liens renvoient vers les fiches correspondantes de notre [base bibliographique collaborative](#), qui comprennent les références complètes et, le cas échéant, des accès aux articles cités (en accès libre ou en accès payant, selon les cas et selon les abonnements électroniques souscrits par votre institution) ;*
- *sauf indication contraire, toutes les traductions comprises dans ce Dossier d'actualité ont été réalisées par le rédacteur ;*
- *vous pouvez faire part de vos réactions à ce Dossier en laissant un commentaire sous l'[article](#) correspondant dans notre blog : « Écrans de veille en éducation ».*

Les universités dans les nouvelles vagues de l'évaluation

Un diktat de la mondialisation ?

Pour beaucoup de profanes, la question de l'évaluation de la production des universités a émergé à travers les résultats du « [palmarès de Shanghai](#) ». Il semble que les bourrasques de la mondialisation ou de la [globalisation de l'enseignement supérieur](#) ont imposé à nos gouvernements, à nos universités et à nos universitaires les impératifs de l'évaluation, nouvel avatar du « nouveau management public » (Rey, [2007](#)).

On a alors découvert ou redécouvert que, dans certains pays ou certaines aires régionales, les palmarès de départements de recherche ou d'universités sont largement intégrés dans le fonctionnement ordinaire de l'enseignement supérieur.

Élaborés généralement par des médias ou des organismes privés, les classements et palmarès se présentent ainsi de prime abord comme une source d'information pour le grand public (étudiants notamment) mais sont utilisés, d'une façon plus ou moins avouée, par les pouvoirs publics et le monde des entreprises et ne sont pas sans effet sur les financements et attributions de crédits.

Si certains classements, comme le très médiatisé « [palmarès de Shanghai](#) », sont aisés à critiquer au regard de leurs critères (nombre de prix Nobel par exemple), d'autres grilles d'analyse sont élaborées à partir d'une démarche rigoureuse, à l'image du [CHE University Ranking](#), produit par le *Center for Higher Education*, en Allemagne.

L'évaluation des activités scientifiques, au moins au niveau individuel, n'est pourtant pas une nouveauté dans le monde universitaire. Comme le constate David Bridges (2008), la plupart des chercheurs se sont habitués à la récurrence des procédures d'évaluation scientifique, malgré leurs lacunes et leur faible validité, en partie parce que les instances et les lieux d'évaluation sont divers et que leur carrière comme leurs activités ne dépendent pas d'une seule source de jugement. C'est ainsi qu'une communication refusée dans tel colloque pourra être acceptée dans un autre ou se transformer en un article accepté dans une revue ; qu'un projet non financé par tel appel d'offres pourra trouver des subventions dans le cadre d'un autre programme. Le fait qu'il existe des alternatives en cas de refus, que le champ académique accepte une diversité de critères de jugement, est essentiel aux yeux des universitaires. Quand, en revanche, un seul mode d'évaluation oriente l'essentiel des financements, il devient un outil puissant aux mains des décideurs de politique publique et guide, de fait, la construction de la recherche universitaire et suscite l'inquiétude des chercheurs.

Or, compte tenu de la stagnation des fonds publics alloués à la recherche, de nombreux pays se demandent s'il vaut mieux répartir les crédits de façon équitable à travers tout le système, au risque d'opérer par saupoudrage, ou s'il vaut mieux concentrer les moyens sur quelques départements universitaires pour garantir une plus grande efficacité de la dépense. Dans ce dernier cas, se pose évidemment la question du choix du meilleur indicateur pour orienter cette distribution sélective des crédits.

Les dernières décennies témoignent donc d'une compétition croissante entre chercheurs, universités et revues scientifiques pour obtenir les classements les plus élevés, obsession qui prend le pas sur le traitement des questions les plus cruciales pour la société (Adler & Harzing, 2009). C'est le désir d'identifier des gagnants et des perdants au jeu du prestige académique qui guide le recours aux classements, regrettent ces auteurs, plus que la patiente production des savoirs les plus pertinents.

Une délégation du Sénat français s'est penchée sur la question et a produit en juillet 2008 un volumineux rapport d'information sur *Le défi des classements dans l'enseignement supérieur* (Bourdin, 2008). Le rapport conclut notamment en pointant les lacunes de l'information sur le système et les établissements d'enseignement supérieur, contribuant ainsi à « focaliser l'attention sur des classements du type de celui de Shanghai qui, pour intéressants qu'ils soient, ne reflètent que partiellement et imparfaitement la réalité ».

Comme de nombreux auteurs le reconnaissent, « l'intensification de la bataille autour de la production et de la diffusion des connaissances a pour effet d'accroître le succès et la notoriété des palmarès et autres classements » (Hazelkorn, 2007) des universités et institutions de recherche, quelles que soient les critiques dont ils font l'objet.

Malgré les critiques « rationnelles » de ces classements sur lesquelles s'accordent les responsables universitaires, les établissements dont le classement s'améliore s'empressent de valoriser cette « reconnaissance », quelle que soit sa douteuse valeur académique et, ce faisant, renforcent l'impact des palmarès.

Certes, tous les classements n'utilisent pas les mêmes critères et coefficients de pondération (importance plus ou moins grande apportée à la recherche, l'enseignement, etc.), mais une étude comparative sur des pays anglophones a constaté que les divers classements définissent finalement la notion de qualité en des termes sensiblement identiques (Dill & Soo, 2005), avec une prise en compte prépondérante des productions de la recherche.

Dans ce contexte, le nombre de publications et les indices de citation sont utilisés dans la plupart de ces classements : indices ISI-Web pour le classement de Shanghai, Scopus pour les classements du [Times Higher Education](#) et Google Scholar pour [Webometrics](#), par exemple.

Qualifiés parfois de façon plus ou moins rapides de « bibliométriques », ces indicateurs d'évaluation mettent au centre la question des revues académiques et des publications des chercheurs pour en évaluer la portée scientifique, ce qui explique la sensibilité extrême du milieu universitaire à tout ce qui s'apparente à la ratification institutionnelle des classements de revues. Classer une revue en « A » « B » ou « C », c'est inférer un classement des articles publiés dans cette revue et donc du chercheur qui les a produit...

Comment l'évaluation a saisi les universités françaises

Il serait cependant inexact de conclure que c'est par un détour « international » que le thème de l'évaluation s'est imposé dans le débat universitaire français. Certes, la polémique sur l'évaluation se déploie avec véhémence.

mence depuis le lancement de la [loi relative aux libertés et responsabilités des universités](#) (LRU) et l'installation concomitante de l'[Agence d'évaluation de la recherche et de l'enseignement supérieur](#) (AERES). Mais la question de l'évaluation a été introduite dès les années 80, d'abord dans le débat universitaire (Mérindol, [2008](#)) puis à travers la création du Comité national d'évaluation des universités en 2004.

Selon Jean-Yves Mérindol, l'évaluation est longtemps resté le grand thème absent de tous les débats, du colloque de Caen (1956) à celui d'Amiens (1968) en passant par les réformes qui ont suivi Mai 68. À cette époque, l'évaluation des enseignants-chercheurs s'effectuait essentiellement dans un cadre individuel, lors des procédures de recrutement ou de promotion, à la notable exception de l'évaluation par le CNRS de ses « unités associées » créées en 1965 dans les universités (devenues depuis « unités mixtes de recherche » ou UMR).

Au début des années 80, Laurent Schwartz, mathématicien célèbre, promeut l'idée d'évaluation des universités d'abord dans le « rapport du bilan » demandé par le Premier ministre Pierre Mauroy puis surtout à travers un livre manifeste paru en 1983 sous le titre *Pour sauver l'université*. Inspirateur de la création du Comité national d'évaluation des universités (CNE, dont on peut [consulter les archives web](#)), il en devient le premier président en 1985. Deux particularités fortes du CNE, qui ne font pas alors l'unanimité, tiennent d'une part au caractère indépendant de cette autorité administrative, placée auprès de la Présidence de la République, et d'autre part à sa capacité de publier des rapports publics plutôt que des « rapports au ministre », selon la tradition administrative française.

Parallèlement, Pierre Bourdieu est chargé de remettre les propositions du Collège de France sur l'enseignement à François Mitterrand, Président de la République. Publié en 1985, ce rapport évoque lui aussi la nécessité de laisser une grande autonomie aux universités (avec la gestion d'un budget global), l'évaluation étant la condition d'une concurrence régulée entre les universités, souhaitée par le sociologue.

Afin d'évaluer la productivité scientifique de l'hexagone, l'[Observatoire des Sciences et Techniques](#) est créé en 1990, sous forme de Groupement d'intérêt public réunissant les acteurs de la recherche publique (ministères et organismes de recherche). Il produit des études quantitatives (qui n'existaient pas jusqu'alors en France), appuyées essentiellement sur l'exploitation des banques de données de l'ISI-Web of Science, commercialisé par la société Thomson-Reuters (cf. *supra*), ce qui explique que son champ d'action n'intègre pas les sciences humaines et sociales.

Du côté du CNRS, notons que, depuis plusieurs années, il est demandé aux laboratoires de fournir les indices de citation de leurs chercheurs, calculés à l'aide du logiciel *Publish or perish* (cf. *supra*). Ceci témoigne du recours de plus en plus fréquent aux indicateurs pour produire des appréciations de la qualité scientifique qui soient les plus objectives mais aussi les plus rapides et économiques possibles.

Dans ce contexte, la mise en place en 2007 de la nouvelle Agence d'évaluation (AERES), qui centralise des évaluations autrefois effectuées séparément (évaluation des établissements, évaluation des équipes de recherche, évaluation des formations et diplômes...), a relancé les polémiques sur la façon la plus pertinente d'évaluer les activités universitaires.

La décision de doter une Agence unique de missions d'évaluation portant à la fois sur la production scientifique, les équipes de recherche et les stratégies d'établissement fait aussi débat. « *La concentration de l'évaluation dans une unique agence est contestable car elle établit implicitement une confusion entre des missions qu'il est indispensable de distinguer* », estime par exemple Jean-François Méla, ancien directeur de la Mission scientifique universitaire (Méla, [2008](#)). Le fait de demander à l'AERES une évaluation de tous les établissements sur une période de quatre ans (c'est-à-dire selon le même rythme quadriennal que les contrats d'objectifs entre le ministère et les universités et écoles) fait craindre une industrialisation du processus d'évaluation, peu propice à une compréhension fine des universités (Mérindol, [2008](#)).

Dans le cadre de la contestation de la LRU, particulièrement au sein des universités de lettres et sciences humaines, c'est surtout la question des indicateurs de productivité scientifique qui cristallise les oppositions. Dans un milieu déjà peu à l'aise avec l'exigence [d'indicateurs de performance](#) rendus systématiques par la mise en place de la LOLF (loi organique relative aux lois de finance), la reprise partielle par l'AERES du classement des revues du champ des sciences humaines et sociales réalisé par la Fondation européenne de la science ([ERIH Initial Lists](#)) porte une menace de basculement d'une évaluation essentiellement qualitative par les pairs à une évaluation quantitative sur la base d'indicateurs jugés arbitraires.

□ Et aussi

- CAVET Agnès (2009). « Sur les traces du marché mondial de l'éducation ». *Dossier d'actualité de la VST*, n° 42, février. En ligne : <http://www.inrp.fr/vst/LettreVST/42_fevrier2009.htm>.
- PERELLON Juan-Francisco (2003). *La qualité dans l'enseignement supérieur*. Lausanne : Presses polytechniques et universitaires romandes.

La querelle du classement des revues : l'exemple de la recherche en éducation

Le classement des revues est une question âprement discutée partout dans le monde, même si l'usage bibliométrique des banques de données est plus répandu dans le monde anglo-saxon. De ce point de vue, on peut consulter avec profit les travaux du Centre for the Study of Research Training and Impact de l'université de Newcastle en Australie, qui a constitué et analysé une base de 1 042 revues anglophones en éducation (Fairbairn *et al.*, [2009](#)).

Même en France, à l'instar de ce que fait le CNRS en donnant [un « label » à certaines revues en SHS](#) (hors du champ de l'éducation puisque la discipline n'existe pas dans cet organisme), il est d'usage courant d'établir des listes de revues scientifiques dans lesquelles sont susceptibles de publier les chercheurs. La question est devenue pourtant objet d'actualité plus brûlante depuis un à deux ans, quand l'AERES a voulu explicitement classer les revues pour améliorer l'évaluation des « publiants ».

Du classement de l'ESF (ERIH Initials Lists)...

La Fondation européenne pour la science, en collaboration avec la Commission européenne, a entrepris en 2007 de recenser et de classer les revues par grands domaines de sciences humaines, dans l'objectif avoué de disposer d'un positionnement des « bonnes » revues de recherche européennes dans le contexte international. Il s'agit non seulement de favoriser l'étalonnage de chaque revue au regard des standards de qualité de la communauté scientifique mais aussi de contribuer à dégager des pôles d'excellence européens dans chaque champ. Le projet [European Reference Index for the Humanities](#) (ERIH) a rapidement produit une série de onze listes par domaines disciplinaires, classées en « A », « B » ou « C ». Ce classement n'a pas suscité alors de grandes discussions en France. C'est ainsi que la VST a publié sur son blog en septembre 2007 un [article](#) signalant les résultats du classement dans le domaine de la recherche en éducation (Pedagogical and Educational Research) sans provoquer de réactions.

En théorie, du rang A au rang C, le classement est censé refléter le degré de rayonnement international de la revue. En pratique, il semble que le fait d'être publié en anglais assure une place en A ou en B, indépendamment du niveau de la revue, pendant que la catégorie C est largement réservée aux revues indigènes des pays qui publient dans leur langue nationale. Un constat confirmé, s'il en était besoin, par le label A attribué à de nombreuses revues américaines dont le caractère international est plus que douteux. Il en est ainsi de la [Harvard Educational Review](#), revue de fait réservée à des auteurs américains sélectionnés par un comité éditorial exclusivement américain et qui n'aborde généralement que des sujets concernant l'éducation aux États-Unis..

Des jugements sévères, notamment sur cet aspect du classement, ont été souvent exprimés un peu partout en Europe, y compris dans les pays anglophones, comme en témoigne un rapport de l'Académie anglaise qui conclut que le classement ERIH ne représente actuellement pas une source fiable pour construire des indicateurs métriques des revues à comité de lecture (Weale, [2007](#)).

Concernant le domaine francophone, trois revues sont classées en A :

- *Revue française de pédagogie* ;
- *Histoire de l'éducation* ;
- *Enfance*.

Sept revues sont classées en B :

- *Éducation et sociétés* ;
- *Éducation permanente* ;
- *Formation Emploi* ;
- *Penser l'éducation* ;
- *« Perspectives en éducation »* ;
- *Repères* ;
- *Revue internationale d'éducation*.

Sept revues sont classées en C :

- *Aster* ;
- *Carrefours de l'éducation* ;
- *Didaskalia* ;
- *Éducation et formation* ;
- *L'orientation scolaire et professionnelle* ;
- *Les sciences de l'éducation pour l'ère nouvelle* ;
- *Le Télémaque* ;
- *Recherche et formation*.

On peut déjà remarquer que ne figurent pas dans la liste des revues récentes, mais aussi certaines revues comme *Distances et savoirs* (publiée depuis janvier 2003) ou *Les dossiers des sciences de l'éducation* (publiée depuis 1999). En revanche, la revue *Enfance*, située dans le champ de la psychologie plutôt que des sciences de l'éducation, est l'une des trois revues classées en A, pendant que l'on trouve en B un titre « *Perspectives en éducation* », dont on peut se demander s'il s'agit de la défunte revue *Perspectives documentaires en éducation* de l'INRP dont le dernier numéro est daté de 2005...

...au classement de l'AERES

Il faut attendre l'été 2008 pour que les classements publiés par la Fondation européenne pour la science s'attirent des critiques véhémentes en France, quand l'AERES décide de s'en inspirer pour produire à son tour une « [liste des revues scientifiques du domaine des sciences humaines et sociales](#) ». Dès lors, de nombreuses polémiques, pétitions et prises de position dans diverses disciplines critiquent l'établissement de ces listes, remettant en cause le principe même de classement hiérarchisé et/ou les critères d'inclusion et de classement, jugés insuffisamment clairs (cf. [un exemple avec la science politique](#)).

Dans le domaine de l'éducation, un classement des revues francophones a été publié sous une déclaration commune de l'AERES et de la présidence et vice-présidence de la 70^e section du CNU (sciences de l'éducation).

Ce classement, [consultable en ligne](#), est substantiellement différent de celui de l'ESF.

En rang A, il retranche logiquement *Enfance*, mais promeut *Formation-Emploi*, *Recherche et formation*, la *Revue des Sciences de l'Éducation* et, étrangement, « [Raisons éducatives](#) », qui est une collection de De Boeck (pilotee par la section des sciences de l'éducation de l'université de Genève) et non une revue au sens usuel du terme.

En rang B, il promeut quasiment l'ensemble des revues classées en C par l'ESF et en ajoute d'autres (telles que les *Dossiers des sciences de l'éducation*, *Éducation et didactique* ou *Spirale*), y compris des revues qui n'existent plus (*Année de la recherche en sciences de l'éducation*). En C, ne restent plus que 4 revues, dont *Éducatifs et Formations*, la revue du ministère qui est la seule « dégradée » de B en C.

Enfin, un rang D est créé par l'AERES pour les revues « professionnelles » ou de « vulgarisation », dans lesquelles sont classées non seulement *Sciences humaines*, les *Cahiers pédagogiques* et *Questions vives*, mais aussi la *Revue internationale d'éducation de Sèvres*, éditée par le CIEP, qui était classée en rang B par l'ESF...

Un débat inépuisable ?

La simple confrontation de ces deux classements montre bien la fragilité des critères utilisés, qui apportent parfois plus d'information sur le positionnement des experts à la source des classements que sur le positionnement des revues.

D'ailleurs, Philippe Jeannin était allé jusqu'au bout de cette démarche, en réalisant pour la direction de la recherche une *Revue-métrie de la recherche en sciences humaines et sociales* (Jeannin, 2003). Il s'agissait déjà d'assurer de « *solides fondements à une évaluation de la recherche en SHS publiée dans les revues scientifiques* », en adoptant comme méthode un « tri raisonné » des bases de données utilisées dans chaque discipline pour établir une liste de revues soumises ensuite à l'avis des chercheurs. Ainsi selon lui, « *est scientifique la revue qui est jugée telle par la communauté* ».

Dans le domaine des sciences de l'éducation, le résultat de l'enquête (basée sur 108 réponses de 411 enseignants-chercheurs) donne une image plutôt brouillée de la communauté scientifique.

D'une part, on observe une étrange « hiérarchie » des réponses, où figurent notamment en haut du classement et juste derrière la *Revue française de pédagogie*, la *Revue française de sociologie* et les *Actes de la recherche en sciences sociales*, deux revues de sociologie dont le contenu ne traite que rarement d'éducation. Sur ce qui est présenté comme les « 69 premières revues de Sciences de l'éducation », on dénombre également une dizaine de revues de psychologie, 5 revues de sociologie (non spécialisées sur l'éducation), une revue de science politique ainsi qu'un certain nombre de revues généralistes (*Esprit*) ou au contraire spécialisées (*Pour*, *Communications*, *Langage et Société*...) aux marges de l'éducation.

D'autre part, on constate une méconnaissance générale des revues non francophones : tous les titres anglophones, sauf un, ont une majorité de réponses de type « *ne sais pas* » et certains titres de référence sont étrangement absents (comme la revue *British Educational Research Journal*). Enfin, de façon générale, les titres cités et leur popularité auprès des chercheurs montrent que l'on est plus en présence d'une addition d'écoles, de réseaux ou de sous-champs disciplinaires artificiellement agrégés que d'une communauté scientifique structurée autour de référents communs.

La confrontation des trois classements que l'on vient d'examiner rend par conséquent improbable l'existence d'une liste de revues hiérarchisée qui fasse consensus parmi les chercheurs en éducation. Même la distinction entre revues à comité de lecture et autres revues semble poser problème puisque, selon les classements, on voit bien qu'une revue comme celle du CIEP (*Revue internationale d'éducation*) est considérée comme une revue de bonne qualité (rang B) dans la liste ERIH mais traitée comme une revue de vulgarisation par l'AERES et le CNU !

Pour ceux qui font profession d'évaluer la production scientifique des universitaires et des unités de recherche, une liste des revues de référence est pourtant fort utile, non seulement pour rendre l'évaluation aussi objective que possible mais aussi pour la rendre tout simplement réalisable dans des délais décents (c'est-à-dire sans avoir à lire tous les articles pour jauger leur statut scientifique)...

❏ Et aussi

- Biblio SHS, portail d'information scientifique du CNRS en SHS : <<http://biblioshs.inist.fr/>>.
- Évaluation de la recherche en SHS, blog : <<http://evaluation.hypotheses.org/>>.
- JFM's Blog, blog de Jean-François Méla (catégorie « évaluation ») : <<http://jfmela.free.fr/jfmblog/?cat=8>>.
- UMR Spirit, blog sur l'évaluation quantitative de la recherche en sciences sociales : <<http://bibliometrie.wordpress.com/>>.

Évaluation par les pairs vs évaluation par les chiffres

Peut-on et faut-il vraiment « tout lire » ? L'évaluation par les pairs et ses limites

L'évaluation par les pairs, c'est-à-dire dans la lecture de la production scientifique du candidat ou de son institution par des universitaires du même champ disciplinaire, constitue le mode traditionnel d'évaluation, notamment en SHS. Ce processus s'apparente à celui du « comité de lecture » d'une revue scientifique, qui choisit les articles à publier après une lecture « anonyme » des papiers soumis par d'autres collègues chercheurs. Aux yeux des universitaires, ce principe offre une garantie à la fois de qualité et d'autonomie scientifique contre tout autre critère « non scientifique ».

En France, par exemple, le recrutement et la promotion des enseignants-chercheurs sont établis explicitement sur ce type de processus, bien que parfois, des candidats se demandent comment les jurys de recrutement peuvent sérieusement lire en profondeur des dossiers de milliers de pages en quelques jours.

En fait, une grande part du dispositif d'évaluation par les pairs repose sur l'existence préalable d'une communauté scientifique dans laquelle les chercheurs « évaluateurs » connaissent déjà les grandes tendances de la production scientifique dans leur domaine et sont à même de délivrer un jugement rapide sur une production ou une candidature. Il s'agit moins d'apprécier des travaux « inédits » que de jauger de la valeur relative de publications au regard d'un ensemble d'écrits académiques ou d'autres productions comparables.

On comprend aussi que ce dispositif, très adapté à une communauté scientifique disciplinaire et nationale de taille modeste ou bien fortement structurée au niveau international (ex. les physiciens), pose problème quand il s'agit d'apprécier les productions scientifiques d'un domaine disciplinaire faiblement structuré et organisé de façon essentiellement nationale.

Ce principe d'examen approfondi des productions scientifiques par les pairs peut aussi être porté à une échelle plus importante, comme cela a été fait lors des sessions du *Research Assessment Exercise* (RAE), dispositif utilisé pour évaluer, classer et donner une base d'attribution des moyens à l'ensemble des unités de recherche britanniques.

Le RAE impliquait en effet des jurys de plusieurs centaines d'experts, chargés de lire et d'analyser systématiquement les publications et productions des centres de recherche universitaires (Bence & Oppenheim, 2005 ; Evidence Ltd, 2007). Ainsi, en éducation, 2 000 chercheurs de 86 institutions ont été évalués par un jury d'une soixantaine d'experts (panel regroupé en 2008 avec la psychologie et les sciences du sport pour assurer la comparaison des standards de qualité). Près de 8 000 productions (articles, ouvrages, chapitres d'ouvrages, communications, rapports, logiciels éducatifs...) ont ainsi été parcourues.

On comprend facilement que ce travail impose aux évaluateurs une charge énorme, tant pour collecter les données et préparer les dossiers d'évaluation que pour lire et apprécier les productions scientifiques.

Néanmoins, certaines études ont pointé les limites de l'évaluation par les pairs et surtout montré que cette évaluation produit des résultats comparables à ceux obtenus par l'application de critères bibliométriques.

Dans le champ de la science politique, ainsi, des politologues australiens (Butler & Mc Lister, 2009) ont examiné les 4 400 travaux universitaires qui ont été soumis au panel de science politique du RAE 2001. Des 28 128 citations générées (cf. *supra* à propos de la bibliométrie), ils ont conclu que les citations sont de loin... l'élément le plus important pour prédire le résultat de l'évaluation par le RAE ! Le seul autre élément important corrélé avec les résultats était la présence d'un membre du département dans le panel d'évaluation du RAE, qui permet au département en question d'avoir une connaissance et une compréhension plus intime de la meilleure façon de se mettre en valeur pour la procédure d'évaluation.

Selon eux, un modèle d'évaluation basé sur les indicateurs quantitatifs aurait comme avantage d'éviter ce biais, pour des résultats sensiblement proches de ceux de l'évaluation par les pairs, tout en limitant les coûts en temps et en énergie dans les universités.

Un chercheur britannique de science cognitive parvient à la même conclusion en défendant la nécessité de jauger la validité des indicateurs métriques en les confrontant au meilleur critère externe disponible : celui de la revue par les pairs (Harnad, [2008](#)). Il considère que si des indicateurs métriques (certains plus que d'autres d'ailleurs) sont étroitement corrélés aux classements réalisés par les pairs, alors ces indicateurs peuvent être considérés comme des outils suffisamment robustes et indépendants pour juger de la qualité des productions scientifiques.

La base en grandeur réelle du RAE permet par conséquent des comparaisons entre les résultats de l'évaluation par les pairs et ceux qui auraient pu être déduits des indicateurs quantitatifs, par calculs de corrélation (avec analyse de régression). Autrement dit, il suffit de comparer les classements issus du RAE et ceux résultant d'un calcul des citations des chercheurs des équipes. En l'occurrence, diverses études citées par Harnad ont effectivement montré que les classements par les pairs sont fortement corrélés avec les mesures de citation dans tous les domaines, y compris dans ceux pour lesquels l'évaluation par les pairs ne recourrait pas aux citations (certains panels disciplinaires utilisaient déjà les indicateurs de citation).

Ce type d'analyses explique qu'après une grande session d'évaluation parallèle en 2008, où ont été utilisées simultanément l'évaluation par pairs et l'évaluation par indicateurs métriques, il a été décidé que le RAE serait ensuite effectué sur la base des seuls indicateurs métriques, avec, pour les sciences sociales et humaines, l'appoint d'un processus « plus léger » de revue par les pairs. Lancé par l'instance de financement de l'enseignement supérieur en Angleterre (HEFCE), le nouveau dispositif [Research Excellence Framework](#) est en cours de réalisation, sur la base de nombreuses études consultables sur le site web du programme.

Au-delà du cas du RAE, c'est donc la question du passage d'une évaluation par les pairs à une évaluation largement informée ou appuyée sur des indicateurs métriques qui est posée.

La bibliométrie, une évaluation par les nombres ?

On résume souvent sous le terme de « bibliométrie » le recours à des indicateurs quantitatifs pour évaluer une activité scientifique. En fait, la bibliométrie est l'application de méthodes mathématiques et statistiques aux productions de littérature scientifique, mais elle n'est pas conçue, a priori, comme moyen d'évaluation d'un chercheur ou de son équipe. Si l'objectif de la bibliométrie est bien de décrire la science et de mesurer la production de connaissances à travers la littérature scientifique, elle ne se confond pas avec une méthode de jugement et de répartition des moyens de la recherche.

Comme prévient le directeur du Centre interuniversitaire de recherche sur la science et la technologie (CIRST) à Montréal, il ne faut pas confondre évaluation et bibliométrie, même si, comme il le reconnaît « *le succès des indicateurs sauvages (h index par exemple) et des palmarès (Shanghai par exemple) repose sur la loi sociale qui dit qu'il vaut mieux n'importe quel chiffre que pas de chiffre du tout* » ("any number beats no number"). C'est pourquoi, il estime que « *l'intérêt de bien comprendre les propriétés des indicateurs bibliométriques tient au fait qu'il est impossible d'échapper aux évaluations et qu'il est donc plus efficace de critiquer d'abord rigoureusement les indicateurs mal construits, et dont l'utilisation peut engendrer des effets pervers, pour ensuite montrer qu'il est possible de construire des indicateurs utiles et dont la construction est contrôlée* » (Gingras, [2008](#)).

Pour résumer, les indicateurs bibliométriques sont souvent construits en prenant en compte diverses statistiques concernant le nombre de publications et surtout la fréquence et les sources de citation de ces publications. L'idée de base est que le nombre de citations constitue un indice significatif de l'utilité et de la valeur scientifique de l'article ou auteur cité, ce qui peut apparaître un raccourci rapide, mais finalement pas beaucoup plus que le jugement des pairs qui s'appuie lui aussi bien souvent sur la réputation.

Initiées dans le domaine de la psychologie au début du XX^e siècle (Godin, [2006](#)), ces études se sont déployées dans les sciences exactes qui, pour la plupart, utilisent principalement les articles comme mode de diffusion des résultats de la recherche. La prise en compte des articles de revue est rapidement devenue la méthode privilégiée, permettant d'établir le nombre de citations par auteur et, par ricochet, la « valeur » d'une revue. En effet, le nombre de citations par auteur permet d'établir le coefficient d'impact d'une revue, qui témoigne de son audience et de sa capacité à accueillir les auteurs les plus cités.

Plusieurs revues spécialisées traitent régulièrement de ces questions ([Scientometrics](#), [Journal of Information Science](#), [Journal of Infometrics](#)...) et la plupart des dispositifs institutionnels de recherche, du moins en sciences exactes, ont un recours officiel aux indicateurs bibliométriques. Pour une approche globale et récente de cette question, on peut se référer à un rapport assez approfondi de la commission d'évaluation de l'INRIA (Merlet, [2007](#)).

Du monopole du Web of Science à sa contestation

La plupart des dispositifs bibliométriques s'appuient sur des grandes bases de données commerciales, qui, à partir d'une sélection de revues jugées pertinentes, indexent systématiquement l'ensemble des articles avec leurs auteurs et bibliographies (initialement, l'indexation était limitée aux bibliographies, dont on extrayait les auteurs et titres de publications). L'intérêt de ces bases de données est bien évidemment la possibilité de calculer très rapidement le nombre de citations.

Trois principales bases de données sont aujourd'hui disponibles sur Internet.

Le Web of Science (WOS)

La plus ancienne et la plus connue est le [Web of Science \(WOS\)](#), imaginé par Eugène Garfield dans les années 50, qui répondait au désir de disposer d'un index de citations des publications scientifiques adapté à la spécialisation des sujets (ce que ne pouvait pas assurer une indexation humaine classique par des bibliothécaires). L'objectif était de mettre en évidence les réseaux de références scientifiques en recensant les sources citées dans un article mais également les références corrélées, postérieures à sa publication.

Cette base, initialement intitulée Current Contents (CC), est devenue en 1964 la source du Science Citation Index, qui réalise ses extractions à partir des revues indexées dans le CC. L'ensemble est géré commercialement par l'International scientific Institute (ISI), aujourd'hui propriété du groupe Thomson Reuters. La recension du nombre de citations dans les revues, principalement anglo-saxonnes, retenues par le [Science Citation Index](#) s'est imposée progressivement comme une dimension majeure de l'évaluation dans les sciences exactes et a permis au WOS d'acquérir une place quasiment monopolistique dans son champ.

Thomson commercialise également, dans le cadre de l'ISI-web of science le [Science Social Citation Index](#) (SSCI) et l'[Arts and Humanities Citation index](#), dont la portée reste limitée à certains secteurs disciplinaires (ceux dans lesquels l'anglais est la langue de communication notamment, comme la psychologie et l'économie) et surtout aux aires culturelles anglo-saxonnes.

Il faut noter que le SSCI ne prend en compte que les citations concernant les revues qu'il a déjà sélectionnées. Autrement dit, la citation d'un auteur dans une revue qui n'est pas dans le corpus actuel du WOS (par exemple, une revue francophone) n'est tout simplement pas prise en compte, même si cette citation provient d'une revue du WOS à fort « *impact factor* » !

Scopus

Mise en place en 2004 par le groupe Elsevier, la base [Scopus](#) (associée au moteur Scirus) indexe notamment les articles d'environ 15 000 revues depuis 1996 à la seule condition que la revue soit labellisée par une institution scientifique (pas de commission d'évaluation comme pour les Current Contents). Scopus revendique des millions de pages web « scientifiques » et 33 millions de résumés depuis 1966. Elle s'affirme comme un concurrent sérieux du WOS.

La provenance géographique des titres de revues est plus variée, puisque 60% ne sont pas basés aux États-Unis. Elle compte beaucoup plus de titres en SHS (2 850), mais porte sur une période encore limitée de onze ans (Kosmopoulos & Pumain, [2008](#)).

Il est à noter que Scopus publie en ligne un bulletin mensuel, [Research trends](#), qui traite des questions de bibliométrie.

Google Scholar

[Google Scholar](#) (GS), lancé en 2004, est un moteur de recherche spécialisé dans la littérature scientifique.

Il met à disposition un espace de recherche gratuit dans un corpus d'articles, d'ouvrages et de littérature grise en constante expansion. Si le périmètre du corpus reste incertain (ce qui est le principal reproche adressé à GS) et plutôt contemporain (prédominance relative des références publiées après 1990), il moissonne d'ores et déjà de nombreux catalogues académiques, dont pour la France, par exemple, des banques de données de l'INIST-CNRS, du SUDOC ou des archives ouvertes, y compris dans le domaine des sciences sociales (HAL-SHS), et des monographies, grâce aux accords passés pour la numérisation des ouvrages ([Google Books](#)) avec de nombreux éditeurs privés.

Diverses études ont comparé Google Scholar au Web of Sciences ou à Scopus, concluant que Google Scholar permettait la prise en compte d'un nombre beaucoup plus important de travaux scientifiques pour un auteur donné, en particulier du fait de la prise en compte d'actes et communications de colloques, de thèses, de monographies et de chapitres d'ouvrages (Harzing & van der Val, [2008](#)). Effet qui s'avère d'autant plus significatif pour le domaine des SHS, où la littérature scientifique passe par des canaux plus diversifiés que les seules revues académiques à comité de lecture. A. W. Harzing a développé un logiciel librement diffusé sur Internet, [Publish or Perish](#), qui permet de réaliser soi-même un certain nombre de statistiques bibliométriques sur la base de Google Scholar. Le CNRS réclame d'ailleurs à ses UMR divers indicateurs bibliométriques qui peuvent être obtenus via ce logiciel, explicitement cité par l'organisme dans ses documents d'évaluation.

En revanche, il semble que les calculs bibliométriques utilisant Google Scholar n'aboutissent pas à des « classements » relatifs fondamentalement différents de ceux issus des calculs du Web of Sciences ou de Scopus : la valeur ajoutée de GS tient à la quantité de travaux recensés, à la prise en compte de documents issus d'aires linguistiques non anglophones, de champs disciplinaires moins normalisés que les sciences dures, et à l'inclusion de travaux de jeunes chercheurs qui peuvent ainsi voir reconnues des productions encore peu représentées dans les revues incluses dans les bases de citation.

Les limites des indicateurs basés sur les citations

Le système de citations a été souvent critiqué sous de multiples aspects. Par exemple, le nombre total de citations mesure l'impact du travail d'un chercheur mais l'estimation de celui-ci peut être biaisée par un petit nombre d'articles très cités, ce qui n'est pas vraiment représentatif de la totalité du travail scientifique d'un chercheur, et ce d'autant plus si celui-ci est co-auteur sur ces articles.

Plus récemment, un physicien de Californie, Jorge E. Hirsch, a donc proposé d'améliorer la mesure de la contribution scientifique d'un chercheur en mettant en avant l'indice h (« h » comme Hirsch), qui est le nombre d'articles pour un chercheur donné ayant reçu au moins ce nombre de citations. Par exemple, un chercheur avec un h égal à 30 a publié 30 articles qui ont chacun été cités au moins 30 fois. Tous les autres articles de ce chercheur, quel qu'en soit le nombre, ont donc reçu moins de 30 citations. Un h = 0 témoigne de l'absence d'impact scientifique d'un chercheur, ses articles n'étant jamais cités (Ferrand, 2007). L'un des avantages de cet indice est qu'il prend en compte l'impact de l'ensemble des articles d'un chercheur sur la durée de sa carrière, ce qui explique notamment que l'indice h a été intégré, dès 2006, dans les résultats diffusés par Thomson/ISI. Certains auteurs ont déjà pointé les failles de cet indice ou proposé des corrections « techniques » (Costas & Bordons, 2007).

Néanmoins, la principale critique réside aujourd'hui dans la question même de l'utilisation quantitative de la citation aux fins d'évaluation des chercheurs et de la recherche.

L'idée, en effet, d'utiliser le nombre de citations comme indicateur de qualité scientifique est jugée irréaliste, en considération des pratiques réelles de citation constatées : bibliographies arbitraires et lacunaires des auteurs, pratiques biaisées, oublis délibérés, renvois d'ascenseur et citations de complaisance des membres les plus influents de telle ou telle revue, etc. Au point que Denise Pumain et Christine Kosmopoulos ont pu écrire que « la citation est un sport de combat » (Kosmopoulos & Pumain, 2008) !

Il faut également noter un usage paradoxal de la citation, quand une référence est fréquemment citée pour son point de vue provocateur ou paroxystique... ce qui l'amène à acquérir une popularité « métrique » nonobstant son aspect scientifique marginal.

Une certitude est que la mise en place d'un système d'évaluation massivement basé sur les publications qui obéissent à certains standards « internationaux » (autrement dit les publications de langue anglaise commercialisées par de grands éditeurs) oriente inévitablement l'activité des chercheurs vers la production d'articles calibrés pour être acceptés dans ces revues internationales, au détriment d'autres productions (Bridges, 2008).

Mais quand la citation, initialement utilisée comme outil de mesure, devient un objectif en soi, elle ne peut plus servir de mesure : « *Once an indicator is made a target for policy, it starts to lose the information content that originally qualified it to play such a role* » (Evidence Ltd, 2007).

C'est ainsi que les articles et les prises de position se multiplient contre la course à la publication, facteur d'aliénation pour les chercheurs mais aussi facteur de distorsion, voire d'altération de la qualité scientifique (Lawrence, 2008), notamment quand, pour mesurer, l'impact d'un auteur en fonction des statistiques de ses publications, c'est la couverture d'une revue (sa diffusion) qui est prise en compte plutôt que la qualité de l'article (Roth, 2005).

Segalla notait ainsi que « *beaucoup d'articles seraient mieux appréciés, publiés plus rapidement, et peut-être auraient plus d'impact s'ils étaient publiés dans des revues spécialisées. Pourtant, du fait que les revues spécialisées tendent à bénéficier d'un facteur d'impact moins élevé ou sont moins connues, elles sont évitées par les jeunes chercheurs qui essaient de se construire un impressionnant dossier de publications pour leur promotion. C'est une stratégie compréhensible, mais qui ralentit en fin de compte la diffusion des idées dans la littérature de recherche et étouffe le dialogue scientifique* » (Segalla, 2008).

La révolution de la publication scientifique ouverte

Vers l'épuisement du modèle commercial des revues imprimées ?

La fonction des revues scientifiques s'est ainsi transformée, passant d'un média de communication à un instrument de gestion de carrière, observe Kingsley, qui s'interroge sur la pérennité du format « revue » dans les années qui viennent. Avec la généralisation de la communication électronique, que signifie exactement le fait d'être « publié » pour un article qui a pu être téléchargé des centaines de fois en pré-publication ou dont le résumé a déjà été lu des milliers de fois ? Ne faudra-t-il pas un jour compter le nombre de téléchargements d'un article plutôt que son nombre de citations ? Déjà, les chercheurs explorent de plus en plus Internet pour trouver un article particulier plutôt que de feuilleter systématiquement telle ou telle revue « prestigieuse » : ils ne se contentent plus de faire confiance aux revues « reconnues » pour se tenir informés de l'actualité de la recherche dans leur domaine. Les auteurs se focalisent d'ailleurs paradoxalement sur les revues « qui comptent », pendant que, comme lecteurs, ils s'ouvrent à de larges collections d'articles et de revues : comme auteurs, les chercheurs veulent publier plus, pendant que, comme lecteurs, ils veulent lire moins (Kingsley, 2007).

Une logique qui se comprend quand on considère le spectre des éditeurs universitaires, avec d'un côté des sociétés privées qui affichent des profits énormes et sont intégrées dans de grands empires éditoriaux, à l'image

d'Elsevier et, de l'autre côté, une multitude de presses d'universités ou de sociétés savantes. Sur les 1 042 revues en éducation analysées par l'équipe australienne SORTI, 52 % (540) sont publiées par de grosses maisons d'éditions, dont Routledge (11%), Sage (5%) et Springer (4%) ; ce mouvement de concentration touche aussi le domaine des revues sur l'éducation. En ce début du XXI^e siècle, les grands éditeurs tels que Springer, Kluwer, Thomson scientifique, Blackwell et Taylor & Francis représentent 52% du marché global de l'édition scientifique et technique (soit 9,2 milliards de dollars en 2005), estime Steele Colin (2006).

Dans ce contexte, il observe que les profits des éditeurs sont essentiellement basés sur les ventes de *packages* aux bibliothèques universitaires et organismes de recherche (achat groupé quasi-obligatoire d'abonnements de revues). Ce système de « bouquets » permet de vendre, à côté des titres les plus demandés, des publications accessoires, même si ces dernières ne sont pas vraiment lues. La prise en compte du numérique dans les années 90 a conduit en fait à doubler les tarifs des abonnements (prix papier + prix numérique), ce qui, avec la généralisation des « bouquets » a permis aux gros éditeurs d'accaparer une part croissante des budgets des bibliothèques. Le plus souvent, cet accroissement des coûts dans le domaine des sciences et technologies s'est effectué au détriment des dépenses pour les sciences sociales et humaines en général et pour les monographies en particulier. Beaucoup d'auteurs de sciences humaines ont subi par conséquent un déclin progressif des ventes de leurs monographies ainsi qu'un manque de débouchés adéquats pour diffuser leur recherche.

Les enjeux scientifiques de « l'Open Access »

Une solution résiderait dans le recours à la publication électronique mais, selon Colin, les analyses actuelles, montrent que le conservatisme d'une grande part de la communauté universitaire reste le frein essentiel à l'édition en ligne de monographies spécialisées. La monographie imprimée sur papier tient encore une place centrale dans les processus de titularisation et de promotion. Pourtant, les éditeurs qui fournissent des contenus libres de droits sur le web constatent que le libre accès en ligne a une tendance à stimuler les ventes de sa version papier.

Les archives ouvertes institutionnelles d'universités ont finalement un effet potentiellement plus important pour la diffusion scientifique et la valorisation dans les sciences sociales et humaines que pour les sciences et techniques, qui ont déjà un système de distribution bien établi (même si ce dernier génère par ailleurs des coûts élevés).

« Depuis quelques années un bras de fer s'est enclenché entre les éditeurs commerciaux, qui revendent aux chercheurs/auteurs leurs publications dont ils ont acquis les droits, et le mouvement international de l'Open Access (OA), officialisé par l'appel de Budapest en décembre 2001, qui revendique le libre accès à tous les documents scientifiques » (Kosmopoulos & Pumain, 2008, p. 9).

Le mouvement vers la publication scientifique en accès libre de plusieurs universités de pointe (comme Harvard) menace grandement le système lucratif d'édition des revues académiques. Ceci explique en partie, avancent Adler & Harzing (2009), les résistances du monde académique. Certains universitaires, en effet, sont parvenus au sommet de la hiérarchie académique sur des critères issus des indices de citations dans les revues commerciales. Ils sont souvent maintenant responsables ou animateurs des comités éditoriaux de ces revues. Ils ne voient donc pas d'un œil favorable le développement d'un système de publication en libre accès qui mine le marché des revues scientifiques... et le pouvoir académique et financier qui en découle !

Pourtant, les deux chercheuses sont catégoriques : *« aucun classement qui choisit d'ignorer les travaux publiés sur le web ne restera significatif. Aucun universitaire ne pourra rester pertinent en travaillant sur la base de publications qui, même issues des meilleures revues, sont structurellement en retard de deux ou trois ans sur l'état réel de la recherche »* (Adler & Harzing, 2009, p. 6).

Même si les chercheurs en éducation ne sont pas particulièrement à la pointe dans ce domaine, le nombre des travaux en sciences humaines et sociales qui sont numérisés et mis en ligne connaît indéniablement une croissance exponentielle depuis quatre à cinq ans.

Ces travaux sont publiés dans les revues mises en ligne (cf. par exemple via les portails revues.org ou [Cairn.info](http:// Cairn.info) en France), dans les archives ouvertes institutionnelles d'établissements, de disciplines ou de réseaux de recherche ([HAL_SHS](http:// HAL_SHS) par exemple), sur des sites publics (cf. [Persée](http:// Persée)) voire sur des pages personnelles des chercheurs et enseignants-chercheurs.

Si l'on ne connaît pas encore l'issue des luttes, plus ou moins sourdes, qui se mènent entre les grands éditeurs commerciaux et les partisans du libre accès, le succès grandissant des archives ouvertes et de la convergence vers des méta-données communes et transférables (OAI-PMH, protocole respecté par les plus grands catalogues de bibliothèques publiques et universitaires) est indéniable (Kosmopoulos & Pumain, 2008).

La perspective de « moissonneurs » capables d'interroger l'ensemble de ces entrepôts de documents scientifiques rend caduque une bonne partie des arguments contre la bibliométrie telle qu'elle se pratiquait dans l'ISI web (notamment tout ce qui concerne la restriction aux seuls périodiques, qui plus est d'origine anglo-saxonne).

Plusieurs analyses ont par ailleurs démontré que les publications scientifiques en ligne bénéficient d'un meilleur taux de citation scientifique que celles uniquement disponibles dans le circuit fermé des revues payantes (Hajjem, Harnad & Gingras, 2005 ; Brody, Harnad & Carr, 2006). Les articles en accès libre seraient avantagés par

divers facteurs : accès en avance (via les preprint), biais de qualité (les articles de meilleure qualité sont plus susceptibles d'être mis en ligne et ce sont les plus demandés), usage plus spontané (facilité d'accès) et enfin un avantage différentiel (tant que tous les articles ne sont pas en « open access »).

Aujourd'hui, le perfectionnement des outils de calcul et le progrès des logiciels et de la numérisation permettent l'indexation en langage naturel du texte complet d'un document et offrent même de premières pistes d'analyse de corpus. Une des pistes est actuellement explorée par le projet européen EERQI, visant à trouver de nouveaux indicateurs de qualité pour les productions scientifiques (Botte & Vorndran, [2008](#)).

□ Et aussi

- CAMPBELL Phillip (2008). « Escape from the impact factor ». *Ethics in Science and Environmental Politics*, vol. 8, n° 1. En ligne : <<http://www.int-res.com/abstracts/ese/v8/n1/p5-7/>>.
- STEELE Colin, BUTLER Linda & KINGSLEY Danny (2006). « The publishing imperative: the pervasive influence of publication metrics ». *Learned Publishing*, vol. 19, n° 4, octobre, p. 277–290.
- VAN LEEUWEN Thed (2008). « Testing the validity of the Hirsch-index for research assessment purposes ». *Research Evaluation*, vol. 17, n° 2, juin, p. 157–160.
- WATERS Lindsay (2004). *Enemies of Promise: Publishing, Perishing, and the Eclipse of Scholarship*. Chicago : Prickly Paradigm Press.

Comment évaluer la qualité en SHS ?

Quelle que soit la pertinence des statistiques de citations, la transposition des méthodes d'évaluation des productions scientifiques des sciences exactes aux sciences humaines et sociales s'avère problématique, comme on pouvait s'en douter en examinant déjà la question des classements de revues.

Le décalage est d'ailleurs moins important entre les grands continents disciplinaires que seraient les sciences exactes et les sciences humaines et sociales, qu'entre les disciplines et sous-continentes disciplinaires. Ainsi, plusieurs études ont montré que les mathématiques se rapprochent des sciences humaines du point de vue de certaines pratiques, tandis que la psychologie (qui entretient parfois des relations particulières avec l'éducation) ou l'économie se sentent plus à l'aise avec les protocoles des sciences « dures ».

Ceci rappelé, si l'on examine plus spécifiquement les autres sciences humaines et les sciences de l'éducation, on constate des différences majeures qui tiennent autant aux habitudes de publications scientifiques (avec une place moins importante des revues) qu'à la position même de certaines disciplines engagées dans des pratiques sociales, dont les critères de qualité ne peuvent être exclusivement académiques.

Des pratiques de publication plus diversifiées

Une étude a ainsi analysé les pratiques de citation pour mesurer la place des périodique, en sciences exactes (« *Natural Sciences and Engineering* ») et en SHS, à partir du corpus des bases de l'ISI web of knowledge de Thompson de 1981 à 2000 (Larivière, Archambault, Gingras & Vignola-Gagné, [2006](#)). Elle a mis en lumière le fait que la proportion de citations qui concernent des articles de revues est moitié moins importante en SHS (40%) qu'en sciences exactes (82%), avec une moyenne de 44% pour les sciences de l'éducation. Compte-tenu du fait que l'enquête se base sur un corpus composé lui-même exclusivement d'articles de revues (à l'exclusion de monographies, rapports, communications et autre littérature grise), on peut estimer que le nombre de citations vers des revues constitue ici un plafond (maximum) plutôt qu'un plancher, indicateur des pratiques de citation dans le domaine considéré.

Selon Botte ([2007](#)), parmi les 1 700 revues de sciences sociales et humaines (à comparer aux quelques 8 700 revues de sciences exactes) existantes dans l'ISI web of knowledge, la sélection des titres répondant aux mots-clés « *education* », « *special education* » et « *psychology, educational* » compte seulement 153 revues. 66 % d'entre elles sont américaines et 24% anglaises. Seules sept revues ne sont pas en langue anglaise (trois multilingues, trois allemandes, une russe). Une grande part des revues retenues sont liées à la psychologie et à la médecine. Si l'on considère les 20 revues d'éducation qui présentent le meilleur facteur d'impact, seulement 4 ne sont pas consacrées à la psychologie ou à la médecine.

L'étude SORTI avait pour sa part remarqué que sur 1 042 revues en éducation de langue anglaise, 46% sont publiées aux États-Unis et, malgré cela, seules 18,9% (197) sont prises en compte pour le calcul du facteur d'impact ISI en 2007, la plupart relevant de la psychologie de l'éducation.

Du côté de Scopus, on comptait 69 revues françaises de SHS en octobre 2008, mais aucune dans la spécialité éducation. Outre le rôle moins central des revues dans le domaine des SHS, de nombreuses études ont également montré que le corpus utilisé reste pour l'essentiel américain et de langue anglaise.

Une étude australienne remarquait ainsi qu'en gestion, un enseignant-chercheur français de l'université de Bourgogne ([Gérard Charreaux](#)) avait accumulé 30 citations dans les revues intégrées dans la base ISI web alors que Google Scholar l'en crédite de plus de 1 000 (Adler & Harzing, [2009](#)). La seule raison du différentiel tient à l'utilisation du français, critère sans rapport avec l'impact du chercheur dans son champ. De fait,

l'association de Google Scholar avec un logiciel comme [Publish or Perish](#) constitue à l'heure actuelle le meilleur outil bibliométrique existant pour les disciplines de SHS.

Plus globalement, on constate que peu de banques de données bibliographiques sont pertinentes pour un réel travail bibliométrique en SHS et particulièrement en éducation. Pour un recensement des banques existantes et de leurs limites (dont la banque Francis), on peut consulter le rapport réalisé pour le Conseil de recherches en sciences humaines du Canada sur l'utilisation de la bibliométrie dans les sciences sociales et les humanités (Science Metrix, [2004](#)).

Ces limites des méthodes quantitatives usuelles expliquent en partie les motivations d'un projet européen comme EERQI ([European Educational Research Quality Indicators](#)). Porté par l'université de Hambourg, ce projet réunit une vingtaine de partenaires (universités, centres de recherche, associations de chercheurs, éditeurs universitaires et commerciaux), sous la coordination d'Ingrid Gogolin, par ailleurs présidente de l'Association européenne de recherche en éducation (EERA). EERQI vise à définir et mettre en œuvre de nouveaux indicateurs et de nouvelles méthodologies d'évaluation de la qualité des publications de recherche dans le domaine de l'éducation, en vue d'offrir une alternative européenne aux mesures d'impact de citations actuelles (*Social citation index*, par exemple), mal adaptées à la réalité de la recherche en SHS. Bénéficiant d'un financement du 7^e Programme Cadre européen de Recherche et Développement, ce projet illustre la volonté politique de la Commission européenne de traiter ces questions d'évaluation scientifique multilingue, en opposition aux bases d'indexation et aux classements commerciaux.

EERQI devrait déboucher sur la mise en place d'un prototype recourant largement à l'analyse sémantique des textes, permettant l'évaluation des productions de recherche au sens large (revues, monographies, littérature grise, archives ouvertes, documents électroniques...), dans une perspective multilingue (anglais, allemand, français et suédois pour commencer).

La portée d'une recherche ne se limite pas à son impact académique

L'idée même de juger de la qualité d'un travail de recherche en fonction de seules considérations sur leur mode de production ou de publication apparaît également délicat dans les sciences humaines et sociales.

En effet, dans ce champ, il y a peu de collaborations internationales d'équipe à équipe et peu de consensus normatif sur les modalités de publication : l'importance des contextes sociaux et culturels pour le cœur même de ces disciplines implique encore des agendas scientifiques plus locaux et plus diversifiés, qui rendent problématique l'idée d'une standardisation des modes de production (Science Metrix, [2004](#)).

Dans des disciplines liées à un champ professionnel, l'impact ne doit pas être évalué uniquement parmi les chercheurs, mais aussi dans les communautés professionnelles (Adler & Harzing, [2008](#)).

Or, l'impact d'une recherche en éducation peut-être important pour un système éducatif national mais de peu d'intérêt pour une publication internationale, alors qu'à l'inverse, un article qui répond à tous les standards académiques peut avoir un certain retentissement dans une grande revue internationale mais n'avoir aucune retombée sur le système éducatif local : le « niveau international » des recherches si souvent invoqué, n'est pas strictement équivalent à leur importance sociale. Or, qu'est-ce qu'une recherche en éducation qui se détournerait complètement de son inscription dans les enjeux et les pratiques éducatives ?

En Grande-Bretagne, les auteurs d'un rapport pour le Conseil de recherche économique et sociale (Furlong & O'neal, [2005](#)) ont ainsi plaidé pour l'élaboration d'un nouveau cadre d'évaluation pour la recherche en éducation qui est explicitement menée en relation avec les pratiques (« *explicitly conducted in, with, and/or for practice* »).

L'importance ou la portée d'une recherche en éducation peut être appréciée de façons différentes selon qu'il s'adresse à un public d'universitaires (impact scientifique et académique), de praticiens de l'éducation (« utilité sociale ») ou de décideurs politiques et administratifs (Rey, [2006](#)).

En outre, l'importance qu'a pu prendre telle ou telle recherche peut dépendre d'autres facteurs que sa seule qualité scientifique « interne », sachant qu'un effet de mode peut amener une production à être immédiatement valorisée par le grand public ou reprise par des pouvoirs publics pour légitimer une politique, alors que sa valeur scientifique s'appréciera plutôt dans la durée.

Et aussi

- « Towards a bibliometric database for the social sciences and the humanities: a european scoping project ». En ligne : <<http://www.sussex.ac.uk/Units/spru/esf/index.php>> (consulté le 1^{er} juillet 2009).

Conclusion

Il est vraisemblable que, malgré l'importante force de séduction des indicateurs de qualité chiffrés, l'évaluation des productions scientifiques en sciences humaines et sociales restera encore largement liée à la pratique de l'évaluation par les pairs, du moins pour l'évaluation individuelle des chercheurs.

En revanche, la demande croissante d'indicateurs transparents et comparables, en particulier pour l'évaluation des institutions de recherche (équipes, départements, universités...), peu compatible avec un examen systématique et approfondi des productions, amènera vraisemblablement à poser la question de critères simples, rapides et automatisés pour apprécier la quantité et la qualité des activités et productions scientifiques.

De ce point de vue, le défi pour certaines disciplines de sciences humaines et sociales comme les sciences de l'éducation (ou la recherche en éducation) est à la fois symbolique, financier, institutionnel et épistémologique : comment continuer à exister de façon autonome et reconnue en refusant les indicateurs standardisés existants, sinon en proposant d'autres indicateurs d'évaluation, mieux adaptés au champ ?

La querelle du classement des revues est symptomatique en l'occurrence de l'inadéquation des solutions pour l'instant avancées par les institutions publiques, mais aussi de la fragmentation et du peu de maîtrise de ses revues par la communauté scientifique, au moins au niveau francophone : le simple fait de lister et d'ordonner les revues de recherche existantes dans le champ de l'éducation soulève des problèmes délicats et révèle des méconnaissances de l'existant pour le moins troublantes...

Bibliographie

- ADLER Nancy & HARZING Anne-Wil (2009). « When Knowledge Wins: Transcending the Sense and Nonsense of Academic Rankings ». *Academy of Management Learning and Education*, vol. 8, n° 1.
- SCIENCE-METRIX (2004). *L'utilisation de la bibliométrie dans les sciences sociales et les humanités*. Montréal : Conseil de recherches en sciences humaines.
- BENCE Valérie & OPPENHEIM Charles (2005). « The evolution of the UK's Research Assessment Exercise: publications, performance and perceptions ». *Journal of Educational Administration and History*, vol. 37, n° 2, p. 135-155.
- BOTTE Alexander (2007). « Scientometric Approaches to Better Visibility of European Educational Research Publications: a state-of-the-art-report ». *European Educational Research Journal*, vol. 6, n° 3, p. 303-310.
- BOURDIN Joël (2008). *Enseignement supérieur : le défi des classements*. Paris : Sénat.
- BRIDGES David (2008). *Assessing the quality of educational research in higher education in the UK*. European Educational Research Quality Indicators.
- BRODY Tim, HARNAD Stevan & CARR Less (2006). « Earlier Web Usage Statistics as Predictors of Later Citation Impact ». *Journal of the American Society for Information Science and Technology*, vol. 57, n° 8.
- BROWMAN Howard I. & STERGIUO Konstantinos I. (2008). « The use and misuse of bibliometric indices in evaluating scholarly performance: Esep Theme Section ». *Ethics in Science and Environmental Politics*, vol. 8, n° 1.
- BUTLER Linda & McALLISTER Ian (2009). « Metrics or Peer Review? Evaluating the 2001 UK Research Assessment Exercise in Political Science ». *Political Studies Review*, vol. 7, n° 1, p. 3-17.
- COSTAS Rodrigo & BORDONS Maria (2007). « The h-index: Advantages, limitations and its relation with other bibliometric indicators at the micro level ». *Journal of Infometrics*, vol. 1, n° 3, p. 193-203.
- DILL David D. & Soo Maarja (2005). « Academic quality, league tables, and public policy: A cross-national analysis of university ranking systems ». *Higher Education*, vol. 49, n° 4, p. 495-533.
- RESEARCH POLICY COMMITTEE OF UNIVERSITIES (2007). *The use of bibliometrics to measure research quality in UK higher education institutions*. London : Universities UK.
- FAIRBAIRN Hedy, HOLBROOK Allyson, BOURKE Sid *et al.* (2009). « A profile of education journals ». In *AARE 2008 International Educational Research Conference : Conference Papers, Brisbane 2008*. Victoria : Australian Association for Research in Education (AARE).
- FERRAND Ludovic (2007). « L'indice h : une nouvelle mesure pour quantifier l'impact scientifique des chercheurs ». *L'année psychologique*, n° 107, p. 531-536.
- FURLONG John & OANCEA Alis (2005). *Assessing quality in applied and practice-based educational research*. Swindon : Economic and Social Research Council.
- GINGRAS Yves (2008). *La fièvre de l'évaluation de la recherche. Du mauvais usage de faux indicateurs*. Montréal : Centre interuniversitaire de recherche sur la science et la technologie.
- GODIN Benoît (2006). *On the Origins of Bibliometrics, Project on the History and Sociology of S&T Statistics*. Working paper, n° 33. Montréal : Canadian Science and Innovation Indicators Consortium (CSIIC).
- HAJJEM Chawki, HARNAD Stevan & GINGRAS Yves (2005). « Ten-Year Cross-Disciplinary Comparison of the Growth of Open Access and How it Increases Research Citation Impact ». *IEEE Data Engineering Bulletin*, vol. 28, n° 4, p. 39-47.
- HARNAD Stevan (2008). « Validating Research Performance Metrics Against Peer Rankings ». *Ethics in Science and Environmental Politics*, vol. 8, n° 11, p. 103-107.
- HARZING Anne-Wil & VAN DER WAL Ron (2008). « Google Scholar: the democratization of citation analysis? ». *Ethics in Science and Environmental Politics*, vol. 8, n° 1.

- HAZELKORN Ellen (2007). « L'impact du classement des établissements sur la prise de décision dans l'enseignement supérieur ». *Gestion de l'enseignement supérieur*, vol. 19, n° 2.
- JEANNIN Philippe (2003). *Revue métrique de la recherche en sciences humaines et sociales*. Paris : Direction de la recherche.
- KINGSLEY Danny (2007). « The journal is dead, long live the journal ». *On the Horizon*, vol. 15, n° 4, p. 211-221.
- KOSMOPOULOS Christine & PUMAIN Denise (2008). « Révolution numérique et évaluation bibliométrique dans les sciences humaines et sociales ». *Revue européenne des sciences sociales*, vol. XLVI, n° 141, p. 73-86.
- LARIVIÈRE Vincent, ARCHAMBAULT Éric, GINGRAS Yves & VIGNOLA-GAGNÉ Étienne (2006). « The Place of Serials in Referencing Practices: Comparing Natural Sciences and Engineering With Social Sciences and Humanities ». *Journal of the American Society for Information Science and Technology*, vol. 57, n° 8.
- LAWRENCE Peter A. (2008). « Lost in publication: How measurement harms science ». *Ethics in Science and Environmental Politics*, vol. 8, n° 1.
- MÉRINDOL Jean-Yves (2008). « Comment l'évaluation est arrivée dans les universités françaises ». *Revue d'histoire moderne et contemporaine*, vol. 5, n° 55-4bis, p. 7-27.
- MELET Jean-Pierre (dir.) (2007). *Que mesurent les indicateurs bibliométriques ?* Paris : INRIA.
- REY Olivier (2006). « Qu'est-ce qu'une "bonne" recherche en éducation ? ». *Lettre de la VST*, n° 18. En ligne : <<http://www.inrp.fr/vst/LettreVST/mai2006.htm>>.
- REY Olivier (2007). « Les "petits mondes" universitaires dans la globalisation ». *Dossier d'actualité de la VST*, n° 29.
- ROTH Wolff-Michael (2005). « Publish or Stay Behind and Perhaps Perish: Stability of Publication Practices in (Some) Social Sciences ». *Soziale Systeme*, vol. 11, n° 1, p. 129-150.
- SEGALLA Michael (2008). « Publishing in the right place or publishing the right thing: journal targeting and citations' strategies for promotion and tenure committees ». *European Journal of International Management*, vol. 2, n° 2.
- STEELE Colin (2006). « Book to the Future: 21st Century Models for the Scholarly Monograph ». In *Charleston Conference on Issues in Books and Serial Acquisition 2005*. New York : Howarth Press.
- VORNDRAN Angela & BOTTE Alexander (2008). An Analysis and evaluation of existing methods and indicators for quality assessment of scientific publications. *European Educational Research Quality Indicators*.
- WEALE Albert Weale (dir.) (2007). *Peer Review: the challenges for the humanities and social sciences*. London : The British Academy.

Rédacteur : Olivier Rey

Ce *Dossier d'actualité* est une publication mensuelle de la Veille scientifique et technologique.
Institut national de recherche pédagogique,
membre associé de l'université de Lyon. © INRP

- Vous abonner ou vous désabonner <http://www.inrp.fr/vst/LettreVST/Abonnement.htm>
- Nous contacter <http://www.inrp.fr/vst/Contact.php?nouveau=oui>
- Consulter les nouveautés http://www.inrp.fr/vst/Dernieres_MAJ.php

Veille scientifique et technologique

Institut national de recherche pédagogique

19, allée de Fontenay – BP 17424 – 69347 Lyon cedex 07
Tél. : +33 (0)4 72 76 61 00 – Fax. : +33 (0)4 72 76 61 93