

Revue et évaluation scientifique : fabriquer un monde de classements

La compétition internationale pour l'excellence scientifique s'accompagne d'une batterie d'outils d'évaluation toujours plus nombreuse. Durant ces dernières années, la prolifération des classements de revues en tant qu'instruments de politique publique est particulièrement notoire. Ces classements de revues se présentent comme outil d'évaluation à la fois ajusté aux productions en SHS et alternatif tant aux indices bibliométriques qu'au système d'évaluation par les pairs. Prenant au sérieux cette singularité, cet article analyse les conditions de fabrication de trois classements de revues (en Australie, Norvège, et Europe) et documente leurs importantes variations. Il montre en quoi les classements de revues sont susceptibles de faire émerger un monde spécifique de l'évaluation et discute les enjeux particuliers qui le travaillent.¹

Dans un contexte de compétition internationale, l'évaluation systématique des universités, la rationalisation des organisations de recherche, ou encore l'amélioration des performances et de la productivité scientifiques font l'objet d'une attention croissante de la part des gouvernements et des agences publiques de nombreux pays (Auranen & Nieminen, 2010). Cet engouement pour l'évaluation des productions scientifiques se traduit dans une batterie d'instruments toujours plus nombreuse.

Certains de ces instruments se situent immédiatement dans une perspective mondiale et visent à construire des palmarès d'universités comme ceux du *Times Higher Education*, de l'Université de Shanghai, ou encore du futur « classement de Bruxelles » développé avec le financement de l'Union Européenne². Dans ce cadre, la recherche n'est qu'une partie d'une mesure composite incluant divers indices (enseignement, réputation, nombre de prix Nobel, caractère international...). D'autres instruments ont une visée explicitement nationale comme le *Research Assessment Exercise* britannique (David, 2007), les exercices menés par l'association des universités néerlandaises aux Pays-Bas, ou des systèmes propres à l'Allemagne ou l'Italie (Reale, Barbara & Costantini, 2007). Ces instruments sont quasi-exclusivement centrés sur les productions de recherche, bien que leurs modalités concrètes soient diverses (Geuna & Martin, 2003; Louvel & Lange, 2010).

Parmi ces systèmes d'évaluation, centrés sur les activités de recherche dans un périmètre déterminé, certains reposent sur un instrument élémentaire spécifique : un classement de revues, conçu comme une aide à l'évaluation de la quantité et de la qualité des productions scientifiques de diverses institutions, telles des universités, des organismes de recherche ou des laboratoires. Plusieurs listes et classements de revues ont progressivement vu le jour dès les années 1970, notamment en économie (Moore, 1972; Hawkins, Ritter & Walter, 1973), en psychologie (Koulack & Kesselman, 1975), en sociologie (Glenn, 1971) et en science politique (Giles & Wright, 1975). Souvent centrés sur des enjeux académiques, ces classements de revues n'étaient pas exempts de préoccupations gestionnaires, à l'image du *Science citation index* (Wouters, 1999) ou des incitations financières associées à ceux des écoles de commerce.

1. Nous remercions les membres du comité de rédaction pour la pertinence de leurs suggestions qui nous ont permis d'explicitier plusieurs points de ce texte. Souhaitons que l'esprit collectif qui anime cette revue se réincarne prochainement dans une nouvelle formule.

2. Pécresse V., 2008, « Vers un classement de Bruxelles des universités », Ministère de l'Enseignement Supérieur et de la Recherche, Paris, 13 novembre (<http://www.nouvel-leunivsite.gouv.fr/vers-un-classement-de-bruxelles-des-universites.html?affichage=petit>, consultée le 29.09.2011).

Le caractère inédit des classements de revues contemporains réside davantage dans leur prolifération en tant qu'instrument de politique publique. Le Brésil, la France, la Flandre ont récemment adopté cet outil, et il est en vigueur pour certaines disciplines aux Pays-Bas. Une telle profusion invite à ne pas focaliser l'analyse sur un exemple unique, mais au contraire à rendre compte des variations d'un mouvement d'ampleur à partir d'une sélection raisonnée. Nous avons choisi d'étudier ici les cas de deux pays « marginaux » (Australie et Norvège) cherchant explicitement à mesurer leur importance sur le plan mondial en développant un outil pour l'ensemble des disciplines. En effet en 2004, le gouvernement norvégien et la *Norwegian Association of Higher Education Institutions* (UHR) codéfinissaient un modèle d'évaluation de la recherche destiné à répartir les financements publics dans lequel étaient érigées des listes de revues et d'éditeurs de livres. Et en juin 2008, l'*Australian Research Council* (ARC) présentait ses « draft rankings », résultat d'un vaste exercice de consultation destiné à produire un classement des revues. En contrepoint, un troisième cas nous permettra d'illustrer une démarche centrée sur une partie des Sciences Humaines et Sociales (SHS), sur une base transnationale et ayant pour objectif explicite de valoriser des revues dans des langues minoritaires. Six ans après le lancement du projet, l'*European Science Foundation* (ESF) publiait progressivement en 2007 et 2008 les « initials lists » de l'*European Reference Index for the Humanities* (ERIH). À l'étude détaillée de ces trois classements, nous adjoindrons quelques éléments illustratifs issus d'autres cas, situés en position intermédiaire entre valorisation de la marge et volonté d'internationalisation.

Cet article s'inscrit dans une recherche plus vaste sur les classements de revues en SHS : après avoir adopté une perspective diachronique centrée sur l'économie et la gestion (Pontille & Torny, 2010a), puis développé une analyse synchronique à l'échelle internationale (Pontille & Torny, 2010b), nous focaliserons ici l'attention sur la diversité des formes prises par cet outil d'évaluation. Afin de rendre compte des écarts dans les classements de revues, notre analyse sera volontairement restreinte à leurs conditions de fabrique, sans revenir longuement sur les fortes critiques qu'ils ont suscité de la part d'universitaires et de chercheurs qui, d'un cas à l'autre, partagent nombres d'arguments. Nous mettrons principalement l'accent sur quatre opérations fondamentales communes à ces différents processus afin de saisir leur variation : la désignation des instances chargées d'élaborer les classements de revues, le découpage des domaines disciplinaires, la sélection des revues pertinentes et la qualification pour les hiérarchiser.

Entre évaluation par les pairs et bibliométrie : un outil inédit et multiple

Deux motifs explicites caractérisent généralement la production de classements de revues. Premièrement,

l'objectif est de produire des grilles différenciées qui distinguent les « meilleurs » lieux de publication afin d'inciter les chercheurs à y faire paraître leurs travaux. Cette incitation s'appuie très nettement sur un modèle tourné vers l'international – c'est-à-dire les exigences et standards en vigueur dans une partie du monde anglophone – quelle que soit la discipline, les SHS comprises. Deuxièmement, il s'agit de dépasser les biais des bases de données telles le Web of Science ou Scopus³, maintes fois épinglés dans la littérature (Hicks, 1999 ; Katz, 1999 ; Pons-Novell & Tirado-Fabregat, 2010) pour la faible représentativité des revues en SHS, et tout particulièrement les non anglophones. L'élaboration de classements de revues est donc une façon de combler les lacunes de ces bases de données en prenant en considération des lieux de publication jugés pertinents. Ces deux aspects sont dorénavant bien connus, largement discutés et critiqués (Lemerrier, 2010).

Nous souhaitons insister ici sur un troisième motif bien moins repéré et ses conséquences non négligeables sur les formes d'évaluation de la recherche. Très souvent, les promoteurs de classements de revues cherchent à développer un outil d'évaluation qui soit également ajusté aux productions en SHS par son positionnement particulier parmi la gamme de ceux déjà en vigueur.

En effet, jusqu'à une date récente, l'évaluation de la recherche était tendue entre les indicateurs de mesure et le système d'évaluation par les pairs. Quand les premiers favorisent des jugements synthétiques basés sur des données quantifiables (typiquement le nombre d'articles ou le nombre de citations), le second soutient des jugements fondés sur la lecture minutieuse et approfondie d'un petit nombre de productions. Pour les uns, le système d'évaluation par les pairs tient compte de la complexité et de la diversité des productions, les pairs étant conscients de leurs propres biais (Lamont, 2009), alors que pour d'autres, il produit des jugements largement subjectifs et contingents (Campanario, 1998), et qui prennent trop de temps et par conséquent coûtent cher (Butler & McAllister, 2009). Inversement, certains considèrent que les indices bibliométriques et scientométriques sont objectifs et faciles à produire (Moed, Luwel & Nederhof, 2002), alors que pour d'autres ils portent des jugements à distance, de façon automatique, et qui ont peu de rapport avec la réalité des conditions de production (Norris & Oppenheim, 2007).

3. Une motivation semblable a animé l'élaboration du Latindex dans les pays d'Amérique Latine (Giménez-Toledo, Román-Román & Alcain-Partearroyo, 2007).

La prolifération des classements de revues change les termes de ce constant débat. Bien entendu, on peut réduire un classement de revues à un instrument de mesure, estimant que c'est exclusivement un outil destiné à compter les productions. Mais cette position est immédiatement contrebalancée par l'argument faisant valoir que les revues classées sont justement retenues parce qu'elles sont conformes aux exigences du système d'évaluation par les pairs. Pour la plupart des promoteurs de classements de revues, l'enjeu est d'élaborer un instrument alternatif, précisément du fait de sa position intermédiaire entre les indices scientométriques et l'évaluation directe par les pairs.

Prenant au sérieux cette position, nous explorerons les modes d'existence de cette troisième forme d'évaluation émergente. Bien que la formule générique du «classement de revues» prenne son essor aux quatre coins de la planète, différentes versions se développent parallèlement : leurs modalités de production, les valeurs défendues par leurs promoteurs et leurs usagers, aussi bien que leurs formes concrètes, allant de la simple liste alphabétique au classement sophistiqué, sont extrêmement variées. À la manière des indices de mesure qui ont proliféré depuis 2005 (Bollen, Herbert, Hagberg & Chute, 2009) et la popularisation du *h-index*⁴, ils sont susceptibles de constituer un monde spécifique de l'évaluation. L'étude des trois classements de revues proposée ici nous permettra de discuter en conclusion les enjeux particuliers qui travaillent ce monde en émergence.

Produire des données, désigner des experts, consulter les pairs

La production d'un classement de revues repose sur un dispositif récurrent : les instances qui ont besoin d'une liste s'appuient sur le jugement d'autres personnes reconnues compétentes. Les classements de revues ne résultent pas du travail d'une seule personne, mais toujours d'un processus de délégation à un ou des groupes réunis et mandatés à cette fin (Hicks & Wang, 2011). Ce trait caractérise également les trois cas analysés ici, mais les formes concrètes de cette délégation sont toutefois spécifiques à chacun.

En Australie, les critères d'évaluation ayant des conséquences financières ont été adoptés relativement tôt. À partir de 1995, la production en recherche a été utilisée

4. Cet indice est une mesure composite de la production (nombre d'articles) et de la «visibilité» (nombre de citations) d'un chercheur individuel. Bien que son calcul demeure controversé, l'indice étant fortement corrélé au nombre total d'articles produits, il s'est rapidement propagé dans certains milieux scientifiques.

pour répartir le financement de base, et la proportion de fonds ainsi distribués a progressivement augmenté. En conséquence, la production totale des universités, mesurée à partir du nombre d'articles, notamment ceux référencés par le Web of Science, a cru fortement (Butler, 2004). Des critiques ont alors pointé son découplage avec la qualité, et le gouvernement australien a cherché à adopter une procédure d'évaluation qualitative proche du RAE britannique (Butler, 2008). Une nouvelle institution, le *Research Quality Framework* a été créée sous la tutelle de l'ARC pour proposer d'autres critères. Les mesures classiques, telles que le facteur d'impact ou le nombre d'articles, ont été repoussées au profit de deux types d'indices : d'une part, les citations brutes, d'autre part, des classements des lieux de production, que ce soit des articles, ouvrages ou conférences. Un vaste exercice de classement des articles a ainsi été proposé à l'automne 2007, sur la base de listes préétablies par le *Research Evaluation and Policy Project*, un centre de recherche spécialisé dans les politiques d'évaluation de la recherche et qui a développé une expertise pour les SHS (Donovan, 2005). Des groupes disciplinaires, mis en place par ce centre (et le *Department of Education Science and Training* a insisté sur leur indépendance), étaient invités à opérer une large consultation des communautés scientifiques pour produire des classements organisés en quatre catégories (A*, A, B, C), sur lesquelles nous reviendrons.

En Norvège, les évaluations faites par le *Research National Council*, créé en 1993 par la fusion d'organismes plus spécialisés, se sont déroulées dans les années 1990 sur la base du jugement d'experts internationaux, même si, comme souvent, ces jugements ont été ensuite comparés avec des mesures de citation (Aksnes & Taxt 2004). Réalisées sous l'égide du gouvernement norvégien, ces évaluations visaient à améliorer la qualité des productions scientifiques et à favoriser l'internationalisation de la recherche. Cet objectif était également partagé par l'UHR, qui s'est vu confier en 2003 la réalisation de bases bibliographiques exhaustives (Sivertsen, 2008), afin d'encourager la production de recherche dans les universités. En dépit des variations importantes existant entre disciplines, et en particulier entre les SHS et la plupart des autres disciplines, l'UHR a proposé un système unifié fondé sur le recueil des ISSN (revues) et ISBN (ouvrages et chapitres d'ouvrage) considérés comme académiques (UHR, 2005). En faisant recueillir par un institut de recherche en sciences sociales pendant deux ans les lieux de production effective des chercheurs, l'UHR a montré les insuffisances du Web of Science et construit une première base bibliographique, améliorée avec l'aide de la *National Library of Norway*. Un «comité académique», regroupant les représentants disciplinaires de l'UHR, a ensuite produit des critères pour classer les supports de cette base en deux «niveaux» (level 2 et level 1) et proposé un système de points prenant en compte le type de support, le «niveau» et les cosignatures. Cet instrument a été approuvé par le gouvernement norvé-

gien et a commencé à être utilisé en 2006, les bases bibliographiques étant mises à jour chaque année par consultation publique sur le site web de l'UHR.

Créée en 1974, l'ESF est un regroupement de soixante-dix-neuf organisations consacrées à la recherche scientifique dans trente pays européens. Son objectif est de promouvoir la science de haute qualité à un niveau européen à travers le financement de programmes de recherche et en facilitant les formes de coopération scientifique. Suite à une réunion des représentants des organisations membres de l'ESF à Budapest en 2001, le comité de pilotage pour les Humanités a été chargé de compiler des listes de revues. Entre 2003 et 2004, ce comité a demandé aux communautés scientifiques de fournir, à travers un processus de consultation national, la matière première du projet ERIH afin de constituer une première liste de bonnes revues académiques. Le comité de pilotage a ensuite nommé quinze groupes d'experts (composé de 4 à 7 personnes) pour effectuer un certain nombre d'éliminations, d'additions et de corrections et pour classer les revues retenues en trois catégories (A, B, C). L'objectif était d'établir des premières listes à l'automne 2006. Celles-ci ont fait l'objet d'une large consultation européenne, auprès des organisations membres de l'ESF, des communautés scientifiques nationales, des associations professionnelles et de centres de recherches. Début 2007, les listes ont été révisées par les groupes d'experts à la lumière des 4 818 commentaires formulés au cours de la phase de consultation. Après une accréditation par un comité de l'ERIH, les listes ont été progressivement publiées en 2007 et 2008.

Bien qu'elles partageaient un objectif commun et se lançaient dans l'élaboration d'un outil d'évaluation relativement similaire, les trois organisations n'ont pas développé le même mode opératoire. D'une part, la production d'un premier classement de revues s'appuyait sur des collectifs différents. En Australie, l'élaboration de l'instrument a été confiée, après la constitution préalable de listes de revues, à des groupes disciplinaires placés sous l'égide d'un centre de recherche spécialisé dans l'évaluation académique. En Norvège, l'instrument a résulté du travail d'un comité académique unique, regroupant des représentants disciplinaires, qui a produit des critères pour classer les revues et les éditeurs identifiés préalablement dans une base bibliographique par un organisme spécialisé. Le classement de l'ERIH a quant à lui découlé de l'activité de quinze groupes d'experts internationaux, nommés par le comité de pilotage pour les Humanités, qui ont corrigé et complété des listes disciplinaires préétablies par des institutions des organisations membres de l'ESF. D'un cas à l'autre, les collectifs qui ont produit l'outil d'évaluation « classement de revues » étaient différents tant par leur composition (des groupes disciplinaires, un comité unique, des experts internationaux) que par la teneur de leur principal mandat (organiser une large consultation, élaborer des critères de classement d'une base bibliographique, amender et valider des listes existantes).

D'autre part, la production d'un premier classement de revues résultait d'un processus de consultation des communautés scientifiques très variable. Les groupes disciplinaires australiens étaient chargés de pratiquer une large consultation, en particulier des sociétés savantes, alors que les groupes d'experts de l'ERIH ne consultaient que par la voie institutionnelle des membres de l'ESF. De son côté, le comité académique norvégien de l'UHR s'est d'abord appuyé sur des données agrégées, puis a également consulté les membres universitaires pour l'établissement des premières listes. Dans chaque cas, les groupes n'opéraient donc pas seuls, ils intervenaient à l'interface d'une instance gouvernementale ou européenne et des professionnels invités à s'exprimer sur la première ébauche du classement. Leur action prenait cependant place dans un agencement particulier : les groupes disciplinaires australiens ont consulté les communautés scientifiques une fois les listes érigées par un centre de recherche ; le comité académique norvégien est intervenu après la constitution d'une base bibliographique par un organisme spécialisé et son amélioration par les membres de la bibliothèque nationale ; les groupes d'experts de l'ERIH ont d'abord corrigé la première ébauche des listes, puis les ont révisées à partir des commentaires formulés lors d'une phase de consultation.

Circonscrire des domaines disciplinaires

La variété de l'instrument d'évaluation est également manifeste dans la définition de son périmètre. Les disciplines concernées par les classements de revues ne représentent qu'une partie des domaines scientifiques dont s'occupent les organisations : alors que l'ESF se concentre sur les Humanités pour l'ERIH, l'ARC considère des indices de qualité tout spécialement adaptés aux sciences sociales, et l'UHR distingue trois groupes de disciplines, les SHS se répartissant dans deux (B et C), à l'exception de la psychologie regroupée avec des disciplines ne publiant quasiment que dans des revues.

Premièrement, la délimitation des ensembles disciplinaires était sensiblement différente d'une organisation à l'autre, comme le montre le tableau 1. Avec sa forte orientation en sciences de la documentation, l'UHR a distingué vingt-cinq disciplines qui ne reprenaient pas nécessairement des découpages disciplinaires usuels, mais s'appuyaient sur les catégories du *Web of Science*. L'ERIH a différencié depuis le début quinze disciplines dans les Humanités, qui incluent des ensembles universitaires classiques (e.g. Philosophie, Anthropologie) et des champs de recherche plus contemporains (e.g. Gender Studies). L'ARC a reconnu dix ensembles disciplinaires pour les SHS, allant de « Education » à « Law » en passant par les « Studies in Human Society ». Ces vastes catégories comprennent entre quatre et neuf « champs de recherche » (Fields of Research – FoR), par exemple pour Economics : Economic Theory, Applied Economics, Econometrics, et Other Economics. Le cas de la psychologie illustre bien à lui seul l'importante

flexibilité des rattachements des disciplines à diverses catégories: relevant des «humanités» pour l'ERIH, classée dans le même groupe que les sciences du vivant à l'UHR, identifiée comme une «science humaine et sociale» par l'AERES, la psychologie est aussi rattachée à la section 27 du CNRS en France, «comportement, cognition, cerveau», aujourd'hui intégrée à l'Institut des sciences biologiques.

Deuxièmement, la gamme des disciplines couvertes était également très variable. Si des disciplines classiques se

retrouvaient dans les nomenclatures produites par les trois organisations, elles n'étaient pas toujours appariées de la même manière. Par exemple l'histoire figurait en «History» à l'UHR et à l'ERIH, mais en «History and Archaeology» à l'ARC. De son côté, la sociologie était isolée à l'UHR, alors qu'elle ne relevait pas des humanités à l'ERIH, et qu'elle était un champ des «Studies in Human Society» à l'ARC. Par ailleurs, certains domaines de recherche n'étaient identifiés que dans un seul classement de revues. Ainsi, les «Gender Studies» n'étaient présentes qu'à l'ERIH, le domaine «History and

UHR 2006	ERIH 2008	ARC 2008
Anthropology and ethnology	Anthropology	
Archaeology	Archaeology	
Architecture	Art, architectural design history	
Art history		
Business administration, finance, management		Commerce, Management
Classical studies	Classical studies	
Criminology		
Drama		
Education	Pedagogical and educational research	Economics Education
Geography, demographics and regional development		
History	Gender studies History	History and Archaeology
	History and philosophy of science	
Humanistic media research		
Law		Law and Legal Studies
Linguistics	Linguistics	Language, Communication and Culture
Library and information science		
Labour research		
Literary research	Literature	
Media and communication		
Music	Music and Musicology	
	Oriental and African studies	
Philosophy	Philosophy	Philosophy and Religious Studies
Political science		
Psychology	Psychology	Psychology and Cognitive Sciences
	Religious studies and theology	
Social economics		
Social work		
Sociology		Studies in Creative Arts and Writing Studies in Human Society Tourism and Services

Tableau 1. La gamme disciplinaire dans trois classements de revues

Philosophy of Specific Fields» figurant uniquement à l'ARC, la «Criminology» et «Drama» seulement à l'UHR. La mise en regard avec d'autres exemples de classements de revues nationaux renforce le constat de l'existence de singularités : ainsi au Brésil⁵ et en France, on trouve respectivement «Éducation physique» et «STAPS» qui relèvent du même champ disciplinaire, absent des trois classements étudiés ici. Les Brésiliens considèrent également des disciplines qui leur sont propres, comme «Sciences sociales appliquées» ou «Santé publique». Ces écarts sont bien entendu le résultat d'histoires institutionnelles nationales, mais marquent également l'état de rapports de forces entre disciplines : en France, les «Gender studies» n'ont pas été reconnues comme domaine de recherche autonome, alors que les «Sciences de l'information et de la communication» ont fini par obtenir l'officialisation de leur propre liste de revues.

Troisièmement, de tels découpages posaient très directement le problème des revues interdisciplinaires et transdisciplinaires, point sur lequel les positions des organisations divergeaient. L'ERIH considérait que ces revues pouvaient relever de plusieurs groupes d'experts : «Journals covering overlapping fields may be found in the lists produced by two or more panels (possibly even with different gradings)» (ERIH, 2007 : 1). Inversement, l'ARC soutenait une vision strictement disciplinaire, demandant aux groupes disciplinaires de ne pas classer les revues relevant majoritairement d'autres «champs de recherche» (FoR) et créant une catégorie spécifique de «multidisciplinary journals»⁶. C'était également l'option retenue par l'UHR en définissant deux champs, «cross-disciplinary social science research» et «cross-disciplinary humanities research», où l'on retrouvait les laissés-pour-compte des choix disciplinaires, comme les «women studies» (UHR, 2005 : section 5.4), le plus important étant aux yeux de l'UHR de savoir que ces recherches appartenaient au groupe C avec ses règles propres.

La variété de ces découpages montre donc à quel point des catégories en apparence évidentes et consensuelles comme «les humanités» ou «les sciences sociales» donnent lieu à des arrangements bien spécifiques d'une organisation à l'autre. Une fois les domaines disciplinaires clairement délimités et les groupes de personnes réunis autour d'un mandat précis, l'élaboration proprement dite des listes de revues pouvait commencer.

Sélectionner les revues

Rappelons que pour ces trois organisations, la production d'un classement spécifique aux SHS s'inscrivait explicitement contre les lacunes et les biais du Web of Science. L'enjeu était de confectionner un instrument d'évaluation ajusté aux nombreuses revues, en particulier non anglophones, dans lesquelles publient les chercheurs européens, australiens et norvégiens. Or la volonté de forger des outils qui englobent davantage de supports de publication se heurte à un problème inhérent à toute

entreprise de ce type : formuler des critères pour définir ce qu'est une revue scientifique.

La référence à des procédures typiques du système d'évaluation par les pairs était constante dans les documents des trois organisations. Par exemple pour l'ERIH :

All journals included must fulfil normal international academic standards, i.e. selection of articles is based on an objective review policy. [...] The journals must fulfil basic publishing standards (i.e. ISSN, timeliness of publication, complete bibliographic information for all cited references, full address information for every author) (ERIH, 2007 : 1, italiques dans la version originale).

L'alignement sur des procédures standardisées a clairement orienté le processus de sélection des revues. L'ARC a ainsi proposé à ses comités des premières listes extrêmement longues et, en consultant ses organisations membres, l'ERIH est arrivé à un résultat d'ordre de grandeur similaire (12 212 revues). Mais l'appel à des références partagées sous l'appellation d'évaluation par les pairs ne suffisait pas toujours pour définir les modalités de sélection des revues. Les groupes d'experts étaient potentiellement tendus entre deux options : classer l'ensemble des supports existants qui déclarent s'aligner sur de telles procédures⁷ ou opérer un tri pertinent dans le cadre défini par chaque organisation.

Face à cette alternative, l'ERIH a explicitement adopté une politique de sélection en limitant partiellement ses classements aux revues européennes et en indiquant aux groupes d'experts de ne retenir que les «bonnes revues scientifiques». En conséquence, moins de la moitié des revues présentes au départ ont été retenues dans les «initial lists» (soit 5 888 revues). L'ARC visait en revanche à classer l'ensemble des revues proposées dans les limites de chaque cadre disciplinaire décrit plus haut et, au terme du processus de consultation, a finalement retenu 10 241 revues pour ses «draft rankings». L'UHR prenait une posture intermédiaire, puisqu'il s'agissait d'établir une vision exhaustive des lieux de production des chercheurs norvégiens, mais sous une double contrainte. D'une part, il fallait éliminer les revues «locales» :

The Academic Committee recommends that publications in publication channels with local

5. <http://qualis.capes.gov.br/webqualis/ConsultaListaCompletaPeriodicos.faces> (consultée le 29.09.2011).

6. Ce traitement spécifique a été largement critiqué, étant perçu comme une relégation de la recherche interdisciplinaire par rapport aux publications disciplinaires.

7. Se présenter comme une revue à comité de lecture s'appuyant sur une évaluation par les pairs est aujourd'hui valorisé dans la quasi totalité des cas. Il existe néanmoins des positions dissidentes, où des revues refusent ses procédures au profit des seuls choix du comité éditorial comme *Medical Hypotheses*. Voir «Editor says no to peer review for controversial journal», *Nature*, 18.03.2010, <http://www.nature.com/news/2010/100318/full/news.2010.132.html>.

authorship not be included in the statistics on academic publishing [...] and which therefore be included in the Ministry's basis for setting the research component of the budgets (UHR, 2005: section 4.4).

Définies par le fait que deux tiers des auteurs appartiennent à la même institution, ces revues étaient exclues parce que l'UHR considérait qu'elles ne reposaient pas sur une évaluation externe et que leur prise en compte entraînerait un détournement de l'instrument en provoquant une inflation de publications à des fins purement budgétaires. D'autre part, il s'agissait de trancher au cas par cas, en s'appuyant sur l'existence de procédures d'évaluation par les pairs (UHR, 2005: section 4.2), entre les revues norvégiennes en SHS qui relevaient strictement du monde académique et celles qui s'adressaient aussi à un public plus large. C'est sur cette base que l'UHR a retenu plus de 6 500 revues.

Bien qu'elles s'appuyaient sur une définition procédurale dont les principales caractéristiques sont érigées en standards internationaux, les organisations ont donc façonné des classements sensiblement différents. Chacune définissait les frontières de l'exhaustivité à partir de ses propres impératifs: un maximum de revues dans chaque domaine disciplinaire à l'ARC, une volonté d'exhaustivité sous contrainte de « non-localité » à l'UHR, des supports de publication exclusivement européens et de bonne qualité à l'ERIH.

Définir et valoriser la qualité des revues

Loin d'être une évidence qui s'imposerait naturellement, la qualité d'une revue se présente davantage comme un composite d'éléments éparés et variés. Plutôt que d'opter pour une solution unique, nécessairement réductrice, les organisations ont œuvré en s'appuyant sur une double définition. D'une part, elles ont proposé une définition intrinsèque des niveaux de leur catégorisation, et notamment de la catégorie la plus haute réservée aux meilleures revues. Elles ont ainsi élaboré une série de descripteurs pour qualifier les revues ou les articles qu'elles contenaient. Aussi, l'ARC définissait de cette manière les revues les plus prestigieuses:

Typically an A* journal would be one of the best in its field or subfield in which to publish and would typically cover the entire field/subfield.

8. Pour le groupe A, il était nécessaire que les revues soient également référencées par le Web of Science.

Virtually all papers they publish will be of a very high quality. These are journals where most of the work is important (it will really shape the field) and where researchers boast about getting accepted. Acceptance rates would typically be low and the editorial board would be dominated by field leaders, including many from top institutions (ARC 2008).

Pour sa part, l'ERIH considérait que la catégorie des meilleures revues regroupait celles qui honoraient d'autres critères:

Journals category A: i.e. high-ranking international publications with a very strong reputation among researchers of the field in different countries, regularly cited all over the world (ERIH, 2007: 1).

Enfin, l'URH adoptait une définition du meilleur niveau de publication, spécifiquement pour les disciplines des groupes B et C⁸, fondée uniquement sur une exigence d'internationalisation des supports:

Level 2 channels must be perceived as the leading publication channels in a wide variety of academic contexts, must publish the most outstanding works by researchers from different countries (UHR, 2005: section 5.2).

Ces différents descripteurs dessinaient le prototype d'une « revue de très grande qualité », même si leur évaluation concrète pose problème, en particulier lorsqu'il s'agit de juger de la « qualité » ou de « l'originalité » des critères nécessairement dépendant des univers de références associés (Lamont, 2009). De tels jugements sont le plus souvent relatifs à un ensemble de revues et non absolus, comme on peut le voir avec les qualifications « one of the best » ou « leading ». Néanmoins, ce jugement relatif peut être substantifié, en particulier sous la forme d'une échelle de réputation. Ainsi, en France, l'AERES a indiqué en 2008 dans le court argumentaire qui accompagnait la liste générale des revues que « la classification A, B, C correspond à l'étendue du rayonnement des revues ». On le voit, une définition intrinsèque des lieux de publication, fondée sur des critères descriptifs aussi fins soient-ils, ne garantit donc absolument pas d'aboutir à un classement de revues identique.

D'autre part, les organisations ont introduit une définition relationnelle de leur classement qui s'est principalement exprimé par la mise en place de quotas de répartition. L'ERIH recommandait « that in category A, only 10 to 25% of the total list should appear; this percentage target will differ from one discipline to another » (ERIH, 2007: 1). Le critère de l'UHR était plus strict puisqu'en aucun cas, les revues classées en « level 2 » ne devaient représenter plus de 20% du total de la production – et non des titres – pour chaque discipline. Pour sa part, l'ARC définissait un cadre encore plus contraignant en donnant une clé de répartition complète où la catégorie

A* contenait 5% du classement total, la catégorie A 15%, la catégorie B 30%, et la catégorie C 50%. L'importance relative des catégories de revues était ainsi conçue de plusieurs façons d'une organisation à l'autre.

Or la coexistence de ces deux définitions n'allait pas de soi. Elle générait une tension qui a très nettement orienté le statut de la dernière catégorie. L'objectif relativement ouvert de l'ARC, qui consistait à lister un maximum de revues, a eu pour conséquence de définir la catégorie C négativement: «Tier C includes quality, peer reviewed, journals that do not meet the criteria of the higher tiers» (ARC 2008). Les revues regroupées dans cette catégorie, qui constitue la moitié de la liste, sont celles qui ne sont ni A*, ni A, ni B. En adoptant une démarche explicitement sélective, l'ERIH considérait au contraire qu'au terme du processus la catégorie C, réservée aux revues «with an important local / regional significance in Europe», contenait de bonnes revues:

The Expert Panels emphasise that high quality research appears in journals throughout all three categories. [...] The "C" category represents the real European added value of ERIH. Unlike the journals in the two other categories, "C" journals listed in ERIH fully reflect the linguistic and cultural diversity of Humanities research production in Europe. For many of them, ERIH offers the first opportunity to gain wide visibility also within less specialised research communities (ERIH FAQ, 2008: "Does the categorisation A, B and C reflect differences in quality?" and "Do the C journals make up some kind of residual category?").

Même si l'ERIH défendait sa catégorie la plus basse, l'effet propre du classement est bien de favoriser des publications dans des revues de «haute qualité». Il en est de même au Brésil pour le classement des publications destiné aux doctorants, dont la structure pyramidale vise à valoriser les deux catégories les plus hautes sur les huit créées (Rocha-e-Silva, 2010). En Norvège, l'objectif explicite de la création de deux niveaux de publication par l'UHR était d'inciter la publication dans la plus haute, et c'est sur la base du taux de publication national en «level 2» que l'instrument est jugé comme une réussite (Gilhus & Sivertsen, 2010).

Un monde de multiples classements

Le processus de sélection des revues jugées pertinentes, c'est-à-dire conformes à certaines exigences académiques, la définition des frontières de l'exhaustivité et les critères érigés pour les classer sur une échelle de valeur différaient donc très nettement d'un cas à l'autre. Dans ces conditions, les trois organisations ont produit des instruments qui se distinguaient aussi bien par la taille (plus de 12 000 revues à l'ARC, contre environ 6 000 à l'ERIH et l'UHR) que par les catégories de classement (deux niveaux à l'UHR, trois classes à l'ERIH, quatre à l'ARC). Enfin,

dernier aspect plus directement lié à leur usage, ces classements de revues étaient également distincts dans leur présentation même: alors que l'UHR donnait accès à une base de données en ligne où les utilisateurs peuvent opérer des tris multicritères, l'ERIH publiait des classements séparés, consultables sous forme de fichiers PDF, pour chacune des quinze disciplines retenues, tandis que l'ARC proposait un tableau unique organisé autour des codes de multiples «champs de recherche» (FoR). Sur le site web des trois organisations, le classement de revues s'appréhendait donc soit comme un objet unique, soit il se déclinait en plusieurs sous-listes⁹.

Chaque pays possède donc son propre classement de revues, qui présente des critères issus d'une histoire institutionnelle particulière. Ces initiatives portent néanmoins des valeurs communes: l'international, la qualité, l'excellence. En affichant des valeurs relativement similaires, orientées vers la promotion des productions en SHS, elles ne font pas qu'élaborer des outils jugés alternatifs aux indices de mesures bibliométriques. Leur prolifération même soulève une question majeure à propos du paysage de l'évaluation: va-t-on vers un monde unifié, recréant, sur une base différente, le monopole du facteur d'impact ayant existé en sciences biomédicales? On peut d'abord noter l'absence remarquable des Etats-Unis dans cette production d'instruments, alors même que la place des revues américaines demeure souvent centrale. Deux facteurs permettent de rendre compte de cette exception: premièrement, l'évaluation des chercheurs américains demeure très fortement indexée sur leur activité d'enseignement et non sur leur activité de recherche (Smith, 2011), celle-ci étant jugée indépendamment par l'obtention de financements sur projets évalués par des pairs (Lamont, 2009). Deuxièmement, le système universitaire américain étant très largement décentralisé, avec des financements locaux publics et privés, la production d'un classement de revues, qu'il soit destiné au financement de ces dernières par la puissance publique ou à la répartition de moyens financiers nationaux sur une base homogène, n'a pas de promoteurs. A contrario, la plupart des autres nations/ensembles partagent bien une logique de financement public limité qu'il s'agit de distribuer. Aussi deux éléments distincts permettent d'esquisser l'évolution de ce monde de multiples classements: d'une part, les évolutions propres de chacun de ces classements, dont nous n'avons jusqu'à présent qu'évoqué la genèse; d'autre part, les usages prescrits de ces dispositifs.

9. En France, l'AERES expérimentait plusieurs formules. En 2008, l'outil était disponible sur un site au format html, et se présentait à la fois par ensembles disciplinaires et une liste générale, résultat de leur agrégation. Lors du processus de révision amorcé en 2009, les listes ont ensuite pris la forme de fichiers PDF séparés les uns des autres (<http://www.aeres-evaluation.fr/Publications/Methodologie-de-l-evaluation/Listes-de-revues-SHS-sciences-humaines-et-sociales>, consultée le 29.09.2011).

Journal Menu

- Journal Home
- Aims & Scope
- Author Guidelines
- Editorial Contacts
- View content online
- Association / Society
- Virtual Issues
- Membership

Sales and Services

- Subscribe / Renew
- Recommend to Library
- Customer Services
- Permissions
- Reprints/Offprints
- Advertising

Manuscript Central now live - please see below for more information

Sociologia Ruralis

Published on behalf of the European Society for Rural Sociology

Edited by:
Henry Buller

ISI Journal Citation Reports® Ranking: 2009: Sociology: 19 / 114
Impact Factor: 1.442

Sociologia Ruralis reflects the diversity of European social-science research on rural areas and related issues. The complexity and diversity of rural problems require multi and interdisciplinary approaches. Over the past 40 years *Sociologia Ruralis* has been an international forum for social scientists engaged in a wide variety of disciplines focusing on social, political and cultural aspects of rural development. *Sociologia Ruralis* covers a wide range of subjects, ranging from farming, natural resources and food systems to rural communities, rural identities and the restructuring of rurality.

News and Announcements

Top ▲

Just Published - Special Issue on Mobilities and Ruralities

Click [here](#) to read the latest Special Issue published by *Sociologia Ruralis*

Ranked "A" in the Sociology category of the Excellence in Research for Australia (ERA) rankings

Figure 1. Les labels de qualité d'une revue

En effet, chacun des trois classements étudiés dans cet article, de même que leurs équivalents dans d'autres pays, sont conçus comme des dispositifs en constante révision. Ces révisions impliquent des participations variables des instances du monde scientifique (chercheurs, universités, sociétés savantes, revues), et ont entraîné des modifications substantielles du contenu des listes, parfois du dispositif lui-même: ainsi, dans le cas australien, les quotas entre les différentes catégories ont été abandonnées par l'ARC et, dans le cas français, suite à mouvement critique d'envergure, la hiérarchie interne a été abandonnée par l'AERES au profit d'une simple liste. Plus récemment, suite aux importantes critiques lors de la révision du classement en 2011, l'ARC a également annoncé l'abandon de la catégorisation (A*, A, B, C) au profit d'une liste ordonnée par fréquence de publication des divers supports⁹. Les opérations de révision n'ont donc pas simplement provoqué des argumentations critiques contre l'activité de classement en général ou tel classement particulier, mais se sont accompagnées de transformations de l'outil d'évaluation lui-même.

L'évolution propre à chaque classement s'opère dans un mouvement transnational grandissant. Il est d'abord

9. <http://www.arc.gov.au/era/faq.htm2012> (consultée le 29.09.2011).

10. Alors que pour l'ERIH, la catégorie C contenait des revues de qualité régionale, pour l'AERES seules les catégories A et B étaient qualifiantes. La valeur des revues notées C était donc équivalente à celles qui ne figuraient pas dans le classement. Pour une analyse détaillée du classement de revues de l'AERES, voir Pontille et Torny (2011).

11. La révision du classement australien au printemps 2011 a été de ce point de vue exemplaire, avec des appels lancés par différentes revues à la participation de leurs lecteurs, afin de maintenir ou d'améliorer leur cotation.

présent du côté des promoteurs de classements: les Norvégiens ont observé ce qu'avait mis en place les Australiens (Sivertsen, 2009) pour éviter de produire un outil favorisant la quantité plutôt que la qualité; des chercheurs danois (Lillkåll, 2009) et suédois (Sandström & Sandström, 2009) ont examiné la reprise, pour leur pays, du modèle norvégien; l'AERES a reproduit le modèle de l'ERIH (Merllié, 2009) tout en modifiant le sens des jugements portés sur les revues¹⁰. Loin d'être une opération isolée et totalement indépendante, chaque production d'un nouveau classement de revues s'inscrit dans un monde déjà peuplé par d'autres, où les promoteurs peuvent faire leur marché avant d'élaborer le leur.

Ensuite, ce mouvement transnational existe également du côté des instances classées, en particulier des revues. Celles-ci n'hésitent pas à adopter des positions stratégiques¹¹, à faire de leur place dans un classement un label de qualité (Pontille & Torny, 2010b), voire à multiplier les formes de label. Comme l'illustre la figure 1, une revue européenne de sociologie exhibe à la fois son facteur d'impact et son rang dans le classement australien. Une telle mise en scène des bonnes places obtenues par une même revue dans divers palmarès est clairement au service de sa promotion en tant que support de qualité, voire d'excellence. Toutefois, l'appui sur plusieurs classements de revues n'est pas seulement au service d'une telle posture «marketing». Les revues peuvent aussi s'en saisir dans une perspective critique pour faire valoir leur présence ou leur bonne position dans tel classement afin de la voir améliorer dans un autre. C'est notamment le cas de plusieurs revues qui étaient bien positionnées dans celui de l'ERIH et mal notées ou absentes de celui de l'AERES.

Enfin, les bibliomètres eux-mêmes participent à ce mouvement en comparant le contenu de ces différents classements de revues (Hicks & Wang, 2011), tout comme ils comparaient auparavant ceux produits par des entreprises privées (Norris & Oppenheim, 2007), l'enjeu étant dans bien des cas d'aboutir *in fine* à la couverture pertinente des productions des SHS et au classement le mieux ajusté à la « qualité » des revues. Ces classements sont donc devenus des objets publics transnationaux, mobilisables dans diverses situations et comparables pour servir différents objectifs : produire un nouveau classement ici, améliorer une position dans un autre là, évaluer sa pertinence ailleurs.

Néanmoins, leur espace de prescription demeure national (UHR, ARC) ou potentiellement continental (ERIH). En conséquence, les critiques les plus fortes ont pris forme dans les territoires concernés, par anticipation d'effets induits ou par constat des conséquences après les usages de ces grilles dans des évaluations. Ainsi, la défense de la production scientifique en « langue locale » a émergé comme problème mobilisateur. En effet, en combinant un critère intrinsèque d'internationalité et un quota de 20%, l'UHR n'a laissé quasiment aucune place dans sa meilleure catégorie pour des revues en norvégien, ce qui a entraîné des débats parlementaires et une pétition de 223 chercheurs en SHS pour la défense de leur langue de travail et d'expression, posant directement la question de l'avenir du norvégien comme langue académique (Brock-Utne, 2007). On a pu observer un mouvement similaire de défense des publications en français face au classement de revues de l'AERES. Et si les revues australiennes publient en anglais, elles n'en sont pas moins menacées par l'hégémonie des revues britanniques et américaines (Gläser, 2004), très présentes dans les catégories les plus valorisées du classement de l'ARC. Au-delà de la langue d'expression, la question de l'ajustement par rapport à un standard scientifique purgé de considérations locales ou nationales est clairement en jeu.

Les tensions autour de cet ajustement sont les plus visibles pour une discipline conservant une tradition très particulière d'*authorship* et de publication : le droit (Pontille, 2004). Publiant dans divers supports et très fortement alignés sur les espaces définis par leur droit, les chercheurs de cette discipline n'ont pas pour référence l'international – à l'exception du droit comparé ou du droit international. Aussi, lorsqu'on applique à leurs revues les mêmes critères que pour les autres disciplines, les différents objectifs ne sont pas tenables simultanément. Dans le cas norvégien, valoriser des revues « leaders » revient à renoncer au critère de diversité de l'origine des auteurs puisque 99% d'entre eux sont américains (UHR, 2005 : section 5.5). Dans le cas de l'ARC, les problèmes concrets auxquels peuvent répondre les juristes australiens et qui fondent leur utilité sociale ne peuvent pas figurer dans les articles publiés dans des revues anglaises ou américaines (Svantesson, 2009). En favorisant ces dernières, les classements de l'ARC perdraient de vue leur mission

première. De manière plus générale, c'est la mesure de l'impact sociétal des articles ou productions des SHS qui est discutée, sachant là encore que les mesures classiquement appliquées pour d'autres disciplines – typiquement les brevets – ne sont généralement pas pertinentes. Si les acteurs publics se moquent du lieu de publication et s'intéressent directement au contenu des articles, dont le caractère local et adapté serait alors recherché (Alexander, Lecoutre & Scherer, 2007), d'autres mesures doivent être développées. Une telle vision favoriserait donc la production de classifications locales (thèmes privilégiés, impact sur les politiques publiques), que ce soit à des fins de soutien financier aux revues ou à l'allocation de moyens.

Au final, s'il s'agit de dépasser le Web of Science, celui-ci sert quasi-systématiquement de point de départ et de socle commun à tous les classements, les revues présentes dans cet index étant le plus souvent les mieux classées. Cette configuration particulière fait exister simultanément deux espaces (Hicks & Wang, 2011) : le premier, transnational et commun, le second, national et particulier à chaque pays et liste de revues. Elle découle du fait qu'en tant que dispositif, les classements de revues sont pris dans une double tension. La première réside entre instrument politique et évaluation scientifique, où les classements de revues font l'objet de compromis progressifs entre communautés scientifiques et institutions de gestion de la recherche spécifiques. Une fois ce compromis stabilisé, au moins pour un temps, le dispositif est soumis à une seconde tension entre usages locaux prédéfinis et outil universel, puisque chaque revue, chaque discipline, et chaque institution peut en disposer comme d'une boîte noire pour dénombrer ou qualifier la production scientifique.

David Pontille
pontille@ehess.fr

Didier Torny
torny@ivry.inra.fr

Glossaire

ARC : Australian Research Council

ERIH : European Reference Index for the Humanities

ESF : European Science Foundation

UHR : Norwegian Association of Higher Education Institutions

SHS : Sciences Humaines et Sociales

Références

Aksnes D.W. et Taxt R.E. (2004), « Peer reviews and bibliometric indicators: a comparative study at a Norwegian university », *Research Evaluation*, vol. 13, n° 1, 33-41.

- Alexander J.K., Lecoutre M. et Scherer R.F. (2007), «A global comparison of business journal ranking systems», *Journal of Education for Business*, vol. 82, n° 6, 321-328.
- ARC 2008, «Tiers for the Australian Ranking of Journals».
- Auranen O. et Nieminen M. (2010), «University research funding and publication performance – An international comparison», *Research Policy*, vol. 39, 822-834.
- Bollen J., Herbert, Hagberg A. et Chute R. (2009), «A principal component analysis of 39 scientific impact measures», *PLoS one*, vol. 4, n° 6, e6022-e6022.
- Brock-Utne B. (2007), «Language of instruction and research in higher education in Europe: highlights from the current debate in Norway and Sweden», *International Review of Education*, vol. 53, n° 4, 367-388.
- Butler L. (2004), «What happens when funding is linked to publication counts?», in *Handbook of Quantitative Science and Technology Research*, Dordrecht, Kluwer Academic Publishers, 389-405.
- Butler L. (2008), «Using a balanced approach to bibliometrics: quantitative performance measures in the Australian Research Quality Framework», *Ethics in Science and Environmental Politics*, vol. 8, 83-92.
- Butler L. et McAllister I. (2009), «Metrics or peer review? Evaluating the 2001 UK Research Assessment Exercise in political science», *Political Studies Review*, vol. 7, n° 1, 3-17.
- Campanario J.M. (1998), «Peer review for journals as it stands today – part I», *Science Communication*, vol. 19, n° 3, 181-211.
- David M.E. (2007), «Research quality assessment and the metrication of the social sciences», *European Political Science*, vol. 7, 52-63.
- Donovan C. (2005), «The governance of social science and everyday epistemology», *Public Administration*, vol. 83, n° 3, 597-615.
- ERIH 2007, «Summary guidelines», Septembre 2007, 2 pages.
- Geuna A. et Martin B.R. (2003), «University research evaluation and funding: an international comparison», *Minerva*, vol. 41, p. 277-304.
- Giles M.W. et Wright G.C. (1975), «Political scientists' evaluations of sixty-three journals», *Political Science*, vol. 8, n° 3, 254-256.
- Gilhus N.E. et Sivertsen G. (2010), «Publishing affects funding in neurology», *European Journal of Neurology*, vol. 17, 147-151.
- Giménez-Toledo E., Román-Román A. et Alcain-Partearroyo M.D. (2007), «From experimentation to coordination in the evaluation of Spanish scientific journals in the humanities and social sciences», *Research Evaluation*, vol. 16, n° 2, 137-148.
- Gläser J. (2004), «Why are the most influential books in Australian sociology not necessarily the most highly cited ones?», *Journal of Sociology*, vol. 40, n° 3, 261-282.
- Glenn N.D. (1971), «American sociologists' evaluation of sixty-three journals», *American Sociologist*, vol. 6, n° 4, 298-303.
- Hawkins R.G., Ritter L.S. et Walter I. (1973), «What economists think of their journals», *The Journal of Political Economy*, vol. 81, 1017-1032.
- Hicks D. (1999), «The difficulty of achieving full coverage of international social science literature and the bibliometric consequences», *Scientometrics*, vol. 44, 193-215.
- Hicks D. et Wang J. (2011), «Coverage and overlap of the new social sciences and humanities journal lists», *Journal of the American Society for Information Science and Technology*, vol. 62, n° 2, 284-294.
- Katz J.S. (1999), «Bibliometric indicators and the social sciences», Brighton, E. Sussex.
- Koulack D. et Kesselman H.J. (1975), «Ratings of psychology journals by members of the American Psychological Association», *American Psychologist*, vol. 30, 1049-1053.
- Lamont M. (2009), *How Professors Think: Inside the Curious World of Academic Judgment*, Cambridge, Mass., Harvard University Press.
- Lemercier C. (2010), «Classer les revues de sciences humaines et sociales: comment? pour quoi?», *MethIS*.
- Lillkåll M. (2009), *Evaluation of journal ranking for research evaluation purposes in the Social Sciences*. Master thesis, Royal School of Library and Information Science.
- Louvel S. et Lange S. (2010), «L'évaluation de la recherche: l'exemple de trois pays européens», *Sciences de la Société*, n° 79, 11-26.
- Merlié D. (2009), «La machine de l'AERES à classer les revues de philosophie. Un démontage statistique», *Revue Philosophique de la France et de l'Étranger*, vol. 134, n° 1, 91-99.
- Moed H.F., Luwel M. et Nederhof A.J. (2002), «Towards research performance in the humanities», *Library Trends*, vol. 50, n° 3, 498-520.

Moore W.J. (1972), «The relative quality of economics journals: a suggested rating system», *Western Economic Journal*, vol. 10, n° 2, 156-169.

Norris M. et Oppenheim C. (2007), «Comparing alternatives to the Web of Science for coverage of the social sciences literature», *Journal of Informetrics*, vol. 1, n° 2, 161-169.

Pons-Novell J. et Tirado-Fabregat D. (2010), «Is there life beyond the ISI Journal lists? The international impact of Spanish, Italian, French and German economics journals», *Applied Economics*, vol. 42, n° 6, 689-699.

Pontille D. (2004), *La Signature scientifique. Une sociologie pragmatique de l'attribution*, Paris, CNRS Editions.

Pontille D. et Torny D. (2010a), «Revue qui compte, Revue que l'on compte: produire des classements en économie et gestion», *Revue de la Régulation*, n° 8, [en ligne].

Pontille D. et Torny D. (2010b), «The Controversial policies of journal ratings: evaluating social sciences and humanities», *Research Evaluation*, vol. 19, n° 5, 347-360.

Pontille D. et Torny D. (2011), «Rendre publique l'évaluation des SHS: les controverses sur les listes de revues de l'AERES», *Quaderni*, (à paraître dans le numéro spécial «L'évaluation de la recherche: pour une réouverture des controverses»).

Reale E., Barbara A. et Costantini A. (2007), «Peer review for the evaluation of academic research: lessons from the Italian experience», *Research Evaluation*, vol. 16, n° 3, 216-228.

Rocha-e-Silva M. (2010), «Impact factor, SCImago indexes and the Brazilian journal rating system: where do we go from here?», *Clinics*, vol. 65, n° 4, 351-355.

Sandström U. et Sandström E. (2009), «The field factor: towards a metric for academic institutions», *Research Evaluation*, vol. 18, n° 3, 243-250.

Sivertsen G. (2008), «Experiences with a bibliometric model for performance based funding of research institutions», *Excellence and Emergence. A New Challenge for the Combination of Quantitative and Qualitative Approaches*, Book of Abstracts, 10th International conference on science & technology indicators, Vienna, 126-128.

Sivertsen G. (2009), «A performance indicator based on complete data for the scientific publication output at research institutions», *International Society for Scientometrics and Informetrics Newsletter*, vol. 6, n° 1, 22-28.

Smith H.L. (2011), «L'évaluation de la recherche et des chercheurs en sciences humaines et sociales aux Etats-Unis», in Servais P. (dir.), *L'évaluation de la recherche en sciences humaines et sociales*, Louvain-la-Neuve, Bruylant-Academia, 159-166.

Svantesson D.J.B. (2009), «International ranking of law journals: can it be done and at what cost?», *Legal Studies*, vol. 29, n° 4, 678-691.

UHR 2005, «A Bibliometric Model for Performance-based Budgeting of Research Institutions», Recommendation from the committee appointed by the Norwegian Association of Higher Education Institutions on assignment from the Ministry of Education and Research, 51 p. (uhr.no/documents/Rapport_fra_UHR_prosjektet_4_11_engCJS_endelig_versjon_av_hele_oversettelsen.pdf).

Wouters P. (1999), *The Citation Culture*. Ph.D. Thesis, Amsterdam, University of Amsterdam.