



HAL
open science

The controversial policies of journal ratings: evaluating social sciences and humanities

David Pontille, Didier Torny

► To cite this version:

David Pontille, Didier Torny. The controversial policies of journal ratings: evaluating social sciences and humanities. *Research Evaluation*, 2010, 19 (5), pp.347-360. 10.3152/095820210X12809191250889 . halshs-00568746v2

HAL Id: halshs-00568746

<https://shs.hal.science/halshs-00568746v2>

Submitted on 14 Mar 2011

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

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

The Controversial Policies of Journal Ratings: Evaluating Social Sciences and Humanities

David PONTILLE
IIAC (UMR 8177) CNRS / EHESS
Email: pontille@ehess.fr

Didier TORNY [corresponding author]
RiTME (UR 1323) INRA
Email: torny@ivry.inra.fr

Download printed version:

<http://www.ingentaconnect.com/content/beechn/rev/2010/00000019/00000005/art00004>

David Pontille & Didier Torny, The Controversial Policies of Journal Ratings:
Evaluating Social Sciences and Humanities, *Research Evaluation*, 2010, vol.19(5), p. 347-360
DOI: 10.3152/095820210X12809191250889

Abstract

In a growing number of countries, governments and public agencies seek to systematically assess the scientific outputs of their universities and research institutions. Bibliometrics indicators and peer review are regularly used for this purpose, and their advantages and biases are discussed in a wide range of literature. This article examines how three different national organisations produce journal ratings as an alternative assessment tool, which is particularly targeted for social sciences and humanities. After setting out the organisational context in which these journal ratings emerged, the analysis highlights the main steps of their production, the criticism they received after publication, especially from journals, and the changes made during the ensuing revision process. The particular tensions of a tool designed as both a political instrument and a scientific apparatus are also discussed.

Keywords

Scientific Evaluation • Bibliometrics • Peer review • Quality journals •
Alternative policy tools • Ratings

“The panels assembled for the RAE in the UK are supposed to assess published research. In reality, they assess ability to publish in quality journals, though they make no formal ranking of journals”

(Macdonald and Kam 2007: 711).

Introduction

In a growing number of countries, governments and public agencies are more and more involved in systematic evaluations for assessing the outputs of universities, rationalising research organisations, or augmenting research quality and productivity. Such evaluations serve two main objectives: to improve research performance and to provide tools in decision-making processes about the allocation of research funds (Auranen and Nieminen, 2010).

Over the last thirty years, bibliometric tools, especially those developed by the E. Garfield's Institute for Scientific Information (now owned by Thomson-Reuters) were dominant in the characterisation of evolutions in research. And despite its hesitant beginnings (Wouters, 1999), the impact factor ended up being the main basis for assessment – whether directly or through derived measurements. Two events have recently changed this situation: the development of databases (typically Scopus and Google Scholar) which vie with those of the ISI, and the rapid increase in loosely correlated indices (Bollen et al, 2009; Leydesdorff, 2009) for measuring the importance and quality of scientific productions (Moussa and Touzani, 2010). Faced with this increase in quantitative measurements relating to institutions, researchers, articles and journals, an alternative form of assessment is also being frequently put forward: the peer review, consisting of a group of experts in a particular field, who examine productions such as publications or research projects, in order to judge their quality. This is a principle the origins of which go back to the first faltering steps of modern science, and which provided the basis for the first scholarly journals (Bazerman, 1988; Zuckerman and Merton, 1971).

There is considerable tension between these two forms of assessment. Some people feel that robust quantitative indices provide objective results which are easy to produce and simple to use (Moed et al 2002), whereas judgements based on peer review are generally subjective and contingent (Campanario, 1998), it being a method which takes too long and which is therefore expensive (Butler and McAllister, 2009). Others hold that peer review takes into consideration the complexity and diversity of productions, peers being self-aware of their own biases (Lamont, 2009), and it is clearly favoured by seasoned experts (Brinn et al, 2000). From this point of view, people consider that metrics are based on remote judgements, given in an automatic manner and bearing little relation to the reality of the conditions of production, especially given that databases cover publications in a variable manner (Norris and Oppenheim, 2007), depending on the discipline, and thus leave numerous outlets aside (Nederhof, 2006)¹, particularly those which are not in English (Pons-Novell and Tirado-Fabregat, 2010).

Faced with this polarisation of forms of assessment, some authors have preferred to stress their equivalence, especially within the framework of Research Assessment Exercise in the UK (Charlton and Andras, 2007; Thomas and Watkins, 1998), thus summing up the debate in terms of organisational or financial effectiveness. Conversely, others have pointed out the aporia of this polarisation at the epistemological level by demonstrating that metrics are constantly based upon data shaped by peer review, that peer review could also become remote

¹ A dilemma which echoes the first debates on the value of the impact factor as a proxy for quality (Maier 2006).

(Gläser and Laudel, 2005) or that peer review activities are based on quantified elements (Donovan, 2005; Martin and Irvine, 1983).

In this article we propose to examine a particular mechanism at the heart of these stakes: an cardinal rating of journals (A, B, C) often put forward by research management organisations. These journal ratings are usually designed as a tool adapted to SSH in order to valorise publications in journals – even though books are recognised both by organizations and SSH communities as a significant form of publication (Gläser, 2004), thus shall also be evaluated (White et al, 2009) – by taking into account the linguistic diversity and/or social utility of research. They are also used to assess scientific performances, to organise academic departments into hierarchies (Hix 2004; Kalaitzidakis et al, 1999; Laband, 1985), to discuss the careers and promotions of faculty (Hogler and Gross, 2009), and eventually to modulate the allocation of funds.

Yet despite their proliferation over recent years, journal ratings are rarely analysed head-on, as tools of assessment in their own right – the Spanish case examined by Gimenez-Toledo et al (2007) with its numerous ratings from universities, regional governments and national agencies being an exception. But the production of such lists is not insignificant: each list organises journals into different hierarchical categories and excludes all those which are not on the list. Far from being neutral, journal rating activities always involve stakes which are both cognitive in relation to the categories produced, and political with regard to their effects (Bowker and Star, 1999). By analysing the creation of journal lists, we are here questioning a blind spot in current works on the assessment of academic productions: how are journal ratings produced? Who does them and how? What criteria are used to define categories? How is the division into disciplines managed? What uses are predicted or prohibited? Has any provision been made for a revision process? In what form?

In order to document these questions, we will examine the production of three journal lists in the field of social sciences and humanities: the “draft rankings” produced in June 2008 by the Australian Research Council (ARC), the “initials lists” of the European Reference Index for the Humanities (ERIH) published between 2007 and 2008 by the European Science Foundation (ESF), and the list created by the French Agency for Evaluation Research and Higher Education (AERES) published in July 2008 in France. Our analysis is based on a broad corpus of texts produced by these three organisations in the form of official documents (reports, guidelines) and information provided on their web sites, along with numerous public stances taken (editorials, petitions, points of view, scientific articles, etc.) in relation to these journal ratings².

Above and beyond shedding light on common points, our objective is to characterise variations within a single policy tool for research assessment. After setting out the organisational context in which these three journal ratings emerged, we will analyse the principle modalities of their creation, the criticism they received after publication, especially from journals, and the changes made during the ensuing revision process. Throughout the paper, we will therefore see how the actors perceive journal ratings as an alternative tool to the ISI standard and how they try to raise it to the level of an instrument of assessment which is adapted to SSH and which is an intermediary between peer review and diverse metrics.

² A full list of the sources quoted here is detailed at the end of this paper.

Three organisations in the race for journal ratings

Between 2007 and 2008, organisations as different as ESF, ARC and AERES published lists of journals on their web sites. Their objectives were very similar. They wanted to develop a tool with which to valorise academic productions and to assess the quality of works published in journals which were little or poorly represented in bibliometric databases. And in each case, the disciplines concerned by journal ratings represented only a fraction of the scientific fields covered by these organisations: while ESF concentrated on the Humanities, AERES produced a list for social sciences and humanities and ARC considered quality indices to be specially adapted to social sciences.

The Australian draft rankings

Australian research policy was marked by the very early adoption of assessment criteria with financial consequences. As from 1995, research production was used to distribute basic funding, and the proportion of funds distributed in this manner gradually increased. As a consequence, total university production, measured on the basis of the number of articles – especially those referenced by ISI, grew considerably (Butler, 2004). But this quantitative increase was criticised because it ignored quality, and the Australian government wanted to adopt an assessment procedure similar to that of Britain's RAE (Butler, 2008). A new organisation was created – the Research Quality Framework (RQF) – suggesting the criteria to be used to produce “quality metrics” (DEST, 2006). Classical measurements, such as IF or the number of articles, or more recently webometrics, were set aside in favour of two types of index: citation data on the one hand, research outlet rankings on the other, be they articles, books or conferences. These rankings have to be produced for each discipline, the four-code discipline grid having been defined beforehand in the Australian Standard Research Classification (Castles, 1993).

A vast exercise of article ranking was thus proposed in the autumn of 2007, based on lists which had been pre-established by the Research Evaluation and Policy Project (REPP), a research center which is specialised in evaluation of research policies and had developed an expertise in HSS (Donovan, 2005b). Discipline-based panels, set up by REPP (and the Department of Education Science and Training – DEST - stressed their independence), were asked to organise major consultations with scientific communities in order to come up with these rankings. The names of the members were only made public (web site) after publication of the lists which were presented in a four-tier form (A*, A, B, C).

The ERIH's initial lists

The European Science Foundation is a group of seventy-nine scientific research organisations in thirty European countries³. Its objective is to promote high quality science at a European level, by funding research programmes and by facilitating different forms of scientific cooperation. During a meeting in Budapest in 2001, representatives from the member organisations and researchers debated scientific assessments and, in particular, the low level of representation of the Humanities within the framework of bibliometric tools. They identified two main deficiencies: the low visibility of non-English publications and the diversity of media (conference proceedings, refereed journals, text editions, and so forth) in which humanities research is published (Kritika, 2009). This meeting demonstrated the urgent need for a European Reference Index for the Humanities (ERIH) as an additional tool for research assessment.

³ Further information on membership lists, criteria and procedures: <http://www.esf.org/about-us/79-member-organisations.html> (accessed 22 April 2010).

The ESF's Standing Committee for the Humanities was therefore asked to compile lists of good research journals in fifteen areas of the Humanities. Between 2003 and 2004, the committee asked national panels and scientific communities to provide, through national consultation processes, raw material for the ERIH project, in order to build an initial list of good scientific research journals. The steering committee then appointed fifteen groups of experts (comprised of between 4 and 7 experts, depending on the field) to make "a number of eliminations, additions and corrections" (ERIH ExPanels, 2008: 2) and to classify the selected journals into three categories (A, B, C), the objective being to establish initial draft lists for autumn 2006. These were then the object of broad European consultation with

ESF member organisations, national scientific communities (where judged appropriate by the Panels or the ERIH Steering Committee), subject associations and specialist research centres or advisors (ERIH ExPanels, 2008: 2).

At the beginning of 2007, the draft lists were revised by the Expert Panel members in the light of the 4,818 comments made during this consultation phase. After accreditation by the ERIH Steering Committee, fourteen initial ERIH lists were published incrementally in 2007 and 2008.

The AERES list

Created in 2006 as an independent research assessment agency, AERES had the objective of assessing research laboratories, university courses and research organisations. It defined three sets of disciplines: "hard sciences", "life sciences" and "social sciences and humanities". Like the other two organisations, it tried to acquire tools with which to assess these disciplines and in the autumn of 2007 it set up panels of experts. In order to give journal ratings, these panels sometimes based themselves on pre-existing lists, especially those produced by the CNRS in 2004 to identify journals which deserved its financial backing. We should stress from the outset that as a whole the approach did not benefit from very much publicity, either on the Agency's site or outside⁴.

The expert panels were tasked with classifying journals into three categories (A, B, C), but it was only when work was underway that they were informed of the ERIH's parallel work and that they were given, where pertinent, the lists which were being prepared. Each panel had a discipline manager, relatively autonomous in relation to the Agency, even there is an SSH delegate. After a harmonisation meeting in June 2008, the lists, presented as provisional, were published in July of the same year. It would be more accurate to say that a single list of 6,305 journals, presented as "the list of scientific journals in the field of social sciences and humanities" was made public on the 28th July 2008. In the short argument announcing the publication of this list, AERES thanked everyone who had taken part in the process, but without mentioning any names.

Shaping disciplinary sets

In each case, the production of the first journal rating thus involved the creation of small groups of experts, the names of whom became known (at the very earliest) when the lists were published. These panels did not operate alone however, a scientific community consultation process being part – to varying degrees – of this phase of producing the first ratings. ARC panels used a relatively broad consultation process, including learned societies

⁴ Under conditions of anonymity, we collected the email exchanges of two of these panels, which allowed us to reconstruct the process.

in particular, whereas ERIH panels consulted ESF members via the institutional pathway. Finally, it would seem that AERES practiced no form of organised consultation.

Expert panels were always constituted on a discipline basis, but definition of the disciplines varied from one organisation to another. AERES based itself on fourteen sets of SSH disciplines which are institutionalised in France (e.g. Political Sciences, Economy and Management, Language Sciences). The RQF recognised ten sets of disciplines for SSH; from Education and Law, to Studies in Human Society. These vast categories included between four and nine Fields of Research (FoR); for example, for Economics: Economic Theory, Applied Economics, Econometrics, and Other Economics. From the outset, the ERIH acknowledged fifteen separate disciplines within Humanities, including both standard university disciplines (e.g. Philosophy, Anthropology) and more contemporary fields of research (e.g. Gender Studies). Consequently, the range of disciplines covered varied, with Gender Studies only being listed at the ERIH and History and Philosophy of Specific Fields appearing solely at ARC.

Such categorisation raised the issue of inter-disciplinary and trans-disciplinary journals, and in this respect too, the organisations adopted different stances. The ERIH considered that such journals could fall within the competency of several panels:

Journals covering overlapping fields may be found in the lists produced by two or more panels (possibly even with different gradings)” (ERIH SumGuide, 2007: 1).

Conversely, the RQF opted for a strictly discipline-based approach, asking panels not to classify journals falling by a majority within the scope of other FoRs, and creating a specific “multidisciplinary journals” category. AERES did not give its panels any particular directives, but published an aggregation of lists by selecting the highest score given to each multidisciplinary journal. These differences were found in the final form given to each organisation’s lists. ARC produced a unique ranking based on codes for multiple FoRs, the ERIH created separate ratings for each of the fifteen disciplines, and AERES offered two solutions: a rating by discipline-based group and a general list resulting from their aggregation.

Although they shared the same objective and were trying to develop relatively similar policy tools, the three organisations did not go about it in exactly the same way. In each case, the importance attached to the consultation of scientific communities during the process, the boundaries of the discipline-based groups, and the way in which the journal ratings were presented, were all based on different options. Now that we have described the organisational mechanisms, let us examine the directives given to the panels with regard to the selection and rating of journals within their disciplinary field.

What is a quality journal? How to choose quality journals?

For these three organisations, creating a rating specific to SSH meant extending the list of journals beyond those selected by ISI⁵. Yet the desire to forge tools which covered a greater number of publication outlets hit a problem which is inherent to any rating enterprise of this type: the formulation of criteria with which to define what constitutes a quality scientific

⁵ The same move was made by Latin American countries in order to build the Latindex (Gimenez-Toledo et al, 2007).

journal⁶. Faced with this crucial issue, organisations came up with a threefold definition. First, they placed the emphasis on procedures. At ERIH for example:

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 SumGuide, 2007: 1, originally emphasize).

This reference to the peer review system was a constant in documents from the three organisations and it committed them to a procedure of selecting outlets which conformed to said procedures. The RQF gave its panels extremely long lists. By consulting its member organisations, the ERIH reached a similar number (more than 12,000 journals). The procedural definition of quality journals leads to an alternative: classifying all existing outlets which were in keeping with such procedures or making a pertinent selection within the defined organisational framework. Faced with this alternative, the ERIH clearly adopted a selection policy by partially limiting its ratings to European journals and by directing panels to only retain “good scientific journals”. As a consequence, less than half of the journals were definitively retained from the initial lists (i.e. 5,888 journals). ARC, on the other hand, aimed to classify all journals, following the discipline-based framework described above, and, at the end of the consultation process, retained 10,241 journals for its “draft rankings”.

Second, organisations provided an intrinsic definition of the rating categories, and especially of the highest rating devoted to top quality journals. The categories were constructed by a series of descriptors which qualified the outlets or articles that they contained:

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. 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 Tiers, 2008).

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 SumGuide, 2007: 1).

This list of criteria provided the framework for the prototype for “a very high quality journal”, even if these indices did not stand up to bibliometric analysis (Nisonger, 2002) and even if their concrete assessment was a problem, particularly when it came to judging “quality”, “reputation” or “importance” (Lamont 2009). More often than not, such judgements were relative and not absolute, and it was a definition of this type that AERES chose to differentiate its categories: “the A, B, C rating relates to a journal’s scope of influence” (AERES GenList 2008).

Third, a relational definition of the rating was mainly expressed by the introduction of sharing quotas. Although AERES did not publish any such directives, the report of the Political Science panel shows that they did exist:

The rating was established on the basis of the scores given by the members of the work group and jointly discussed during this group’s meeting on the 24th June 2008. During this discussion,

⁶ A question to which Garfield (1970) himself had to provide an answer in order to ground the perinency of the impact factor.

AERES' management reminded everyone that from a general point of view for all disciplines, journals with A ratings should not represent more than 25% of the total (AERES SciPoList,2008).

For its part, the ERIH recommended "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 SumGuide, 2007: 1). ARC set out an even stricter framework, proposing a formula for division whereby category A* contained 5% of the total ranking, category A 15%, category B 30%, and category C 50%.

The coexistence of the intrinsic and relational definitions of quality journals generated a tension that very clearly orientated the status of the last category of the rating. The effect of ARC's relatively open objective, which consisted in listing as many journals as possible, was to create a negative definition of category C: "Tier C includes quality, peer reviewed, journals that do not meet the criteria of the higher tiers" (ARC Tiers, 2008). The journals in this category, which constituted half of the list, were those which were not A*, or A, or B. Conversely, in adopting an explicitly selective approach, the ERIH considered that when the process was finished, category C, which was reserved for journals "with an important local / regional significance in Europe", contained good journals:

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?").

Finally, the rating produced by AERES could not be separated from its stipulated use: identification of the percentage of "publishing" and "non-publishing" academics in laboratories. To achieve this, one counted books and chapters of books, along with articles published in journals ranked as categories A and B. On the other hand, articles published in journals absent from the rating system or rated as C were not taken into account. The A, B, C rating was thus reduced to a simple Boolean distinction. Such use directly affected the value given to journals: as the AERES rating was partly based on that of the ERIH, the same journal, with a C rating, became an entity without value for AERES.

Above and beyond the general objectives of the assessment and valorisation of SSH productions, analysis of the creation of these ratings could not be separated from the issue of their use. Whilst AERES was explicit about the unique use of its lists, the ERIH was more ambiguous, constructing a classification for the sake of it, the uses of which were multiple:

The ERIH does not encourage using the lists as a basis for calculations in order to assess individual candidates for positions or applicants for research grants. Rather, they may provide a useful tool for aggregate bench-marking of national research systems, for example in order to determine the international standing of the research activity carried out in a given field in a particular country. For individual scholars, the lists will be a useful reference tool when considering where to submit research articles (ERIH FAQ, 2008: "How can the lists of journals be used?").

The ERIH thus considered its device less as a normative tool than as guidelines for professional use. On the other hand, the purpose of the RQF ratings was clearly to directly assess the production of academic departments.

The creation processes for these journal ratings, involving scientific communities in various different forms, did not progress in a linear fashion. The documentary traces produced by the organisations show that cognitive and political tensions left their mark on this rating work. So the AERES panels did not systematically follow an A, B, C coding: for example, the economists preferred to rely on pre-existing rankings and to give distinction to excellent journals by awarding an A* score; the psychologists created a B' category by making the distinction between journals referenced by ISI and those which deserved – according to the experts – such a ranking. The ERIH's philosophy panel emphasised the tension between the ESF member organisations, who were putting forward their own journals, and the “objectivising” criteria they had been given. The Gender Studies panel stated that it was finding it hard to differentiate between A and B journals and that it had finally chosen to use the quota criterion. Some organisations consulted by the RQF panels, such as the Australian Academy of Technological Sciences and Engineering, criticised the logic behind arbitrary quotas which could lead to the creation of very long lists in order to increase the number of A* journals.

Yet retrospectively, these tensions seem very mild compared to the wave of criticism which was to be set in motion by the publication of the ERIH initial lists and of the AERES ratings.

A wave of criticism: the journals strike back

There was some criticism of the ERIH's approach, even before the publication of the initial lists in May 2008: British historians deemed the rating to be “crude and oversimplified” (Times Higher Education 2008a), pointing out that the journals with a C score published articles of central importance at both national and international levels. But after their publication the criticism became far better organised: two editors of history of science journals, Simon Schaffer (British Journal in History of Science) and Iwan Morus (History of Science) contacted colleagues and on the 16th July 2008 produced a manifesto entitled “Journals under threat”. It built a single argument around three distinct elements. Firstly, it considered that the ERIH rating would be used by ESF members to allocate research funding. Secondly, it offered a head-on criticism of the ways in which the rating had been compiled: the lack of consultation with learned societies and with journal editors in the field, the absence of representativeness among panel members and, in the choice criteria, confusion between scope and quality. Thirdly, due to this criticism, the signatory journals asked to be removed from the ERIH ranking. The manifesto was rapidly displayed on numerous academic sites and was published as an editorial by at least 61 journals (Times Higher Education 2009); the fact that most of the journals had received very good scores from the ERIH lent even more force to the demand.

Although criticism of the AERES lists, which were published on the 28th July 2008, was explicitly based on the “Journals under threat” stance, it set out arguments and modes of action which were considerably different. After a few individual stances had been taken, the beginning of October saw the birth of a petition demanding the full and definitive withdrawal of AERES' journal lists. It criticised the murky nature of the chosen criteria, noting their distinct expressions in each individual discipline's arguments: “we don't know if it is a case of measuring “a journal's scope of influence“ (general presentation), its “high level” (criterion retained for sociology) or its “high impact factor” (notion put forward by the psychology discipline)” (AERES Petition 2008). It also pointed out the opacity of the lists, with the names of the experts not being made public, as mentioned above. Finally, it was concerned about the consequences of using these lists, particularly for journals on the fringe (C category or unlisted). Whilst the objective of the petition's promoters was to render the rating device

obsolete, they also envisaged a mode of action similar to that of the British journals, asking editors to demand their removal from the rankings.

The journals' action was based upon two complementary modes. First of all, they published critical articles: in addition to articles on research assessment in general (RHMC 2008), certain analyses related directly to the lists, underlining for example the incoherencies between the different ratings (Briatte, 2008; Merllié, 2009). Secondly, the journals produced editorials which criticised, in particular, AERES' quantitative approach and the absence of any consultative procedure:

Annales does not pretend to suggest – indeed, this is not their role – any ready-made solution to these problems, but can only hope for a joint redefinition of the rules, in a manner which is transparent, autonomous and responsible. (Annales HSS, 2008)

While the world of French research was seeing its greatest ever mobilisation against the numerous reforms underway with unprecedented strikes and demonstrations, the fight against the AERES ratings became symbolic and numerous scientific organisations called for their withdrawal. It was to this end that a text signed by at least forty journals was published on the 8th March 2009, exposing the instrumentalisation of the knowledge published in articles serving a research management that was blind to the scientific stakes which the journals were defending.

The political context also played a major role in the fate of the Australian lists: the RQF became an electoral issue and the newly elected government immediately decided to replace them with a new system, the Excellence in Research for Australia (ERA) initiative. The mechanism for producing the lists, along with the related “four tiers”, was nevertheless maintained. Although criticism of the mechanism remained strong in certain disciplines (notably in Law) and ERIH lists were used as a counter-example of “good rankings” (Hainge, 2008), it was mostly outside of Australia that the journal ratings were criticised – by British philosophers for example (Times Higher Education, 2008b).

The journals therefore adopted public stances, but their criticism and their preferred modes of action were not homogeneous. Analysis of the movements described above allows us to distinguish between four types of reaction to the ratings (cf. Table 1).

Forms of criticism	Judgements of the rating	Modes of action
Radical criticism	To be deleted	Withdrawal of lists (AERES, ARC) Removal of journals (ERIH)
Strong criticism	To be completely redone	General consultations, nomination of experts, establishment of criteria
Equipped criticism	To be revised	To be based on explicit criteria in order to modify the rating
No criticism	To be maintained	Valorisation of the rating, labelling

Table 1. The forms of criticism made by the journals

In contrast with these critical stances, certain journals fully accepted the rating. As with other assessment tools, some journals which had received high scores were in fact able to profit from their ratings. For example, this was the case for the Australian Archaeology journal,

which, in March 2009, proudly published its ranking in various different ratings on its web site⁷:

Australian Archaeology has just been ranked as a 'Tier A' journal by the Australian Research Council (ARC) in the Humanities and Creative Arts (HCA) journal list that will be used to evaluate research as part of the Excellence in Research for Australia (ERA) initiative. This is a welcome addition to the journal's rank in the top band of journals (an 'A' category) in the initial listing of the European Reference Index for the Humanities alongside journals such as *World Archaeology*, *Antiquity* and the *Journal of Archaeological Science*.

In adopting such a stance, journals not only validated the existence of the ratings, they turned them into veritable quality labels aimed at their readerships and at potential authors, as can be seen from their journal information pages⁸.

Between acceptance of the ratings and the different forms of head-on criticism, equipped criticism was aimed at an error of qualification. For example, in the Australian case, the *Educational Philosophy and Theory* journal complained about its B rating. It considered that the organisation which had been consulted – the Australasian Association of Philosophy – was competent for philosophy but could not correctly assess a journal whose subject “is primarily education”, and which was considered to be a “top journal” in a world-wide rating exercise. At the end of the day, it was the issue of separating journals in accordance with the FoR which was questioned and not the legitimacy of rating as an operation (EPAT, 2008). Similarly, several French journals stressed the international dimension of their editorial board, of the authors of published articles, or of the places from which their online articles had been downloaded! (Norris, 2008). So even if the concrete forms still lacked transparency, they based themselves on the values specified by the organisations, such as discipline-based judgement or the journal's international influence. On the other hand, in this critical form, as with acceptance, the question of uses of the list disappears. The rating is then no more than a technical matter – the mutual adjustment of the judgement criteria and the concrete performances of the outlets under assessment.

The stances taken by journals, especially the most critical ones, are an important historical novelty. Indeed, compared to the evaluative instruments previously developed, journals had remained relatively passive or had accompanied the movement, stressing the fact that their quality was fully attested within an audit society approach (Power, 1997). It is only recently that we have seen the appearance of an assumed criticism of the impact factor, or at least, how it is used in assessment (Moed and van Leeuwen, 1996; Seglen, 1997). As far as ratings are concerned, journals were the silent and inert objects to be assessed, rising up as actors who were competent to judge what made scientific “quality” or “high quality”. Whatever the critical form adopted by the journals, individually or jointly, their voice was henceforth part of the rating process. Although none of the organisations had ever consulted or asked their panels to consult them, they were in all cases part of the revision process.

The revision process or the struggle for criteria

As their name suggests, the ratings produced by the organisations were designed to be temporary, and therefore subject to revision. At AERES for example, discipline-based discussion forums were supposed to be opened after the lists were published. At the ERIH,

⁷ <http://www.australianarchaeologicalassociation.com.au/node/3852>

⁸ As examples, *Sociologica Ruralis* (<http://www.wiley.com/bw/journal.asp?ref=0038-0199>) or *Gender, Work & Organization* (<http://www.wiley.com/bw/journal.asp?ref=0968-6673>).

there was a plan to renew the panels and to introduce a feedback procedure for the initial lists. Finally, within the RQF framework, lists were supposed to be submitted for consultation. Generally speaking, there was thus a shift from a relatively narrow process to a relatively broad consultation once the ratings had been published.

The unexpected uprising of the journals significantly modified this consultative broadening. First of all, journals played a central role in revamping the lists - ERA and the ERIH provided editors and publishers with specific forms. The latter thus had the opportunity to describe their journal's activity in a detailed manner, both in terms of core discipline and in quantitative (foreign authors rate, rejection rate, etc.) and qualitative (composition of the editorial board, article assessment procedures, solicited contributions or not). With such descriptions, editors and publishers were able to justify the request that their journal be rated, requalified or removed.

Second, all members of the scientific community were promoted to the rank of potential experts, because they too could intervene, either to describe journals which were not among the first ratings or to argue for a change in score. ERA thus introduced a form aimed at all scholars who wished to put a journal forward for the list. There were just three conditions for such a proposal: that the journal be "a scholarly, peer reviewed journal with an ISSN", that the person making the proposal state whether he/she was a member of the editorial board, and that the decision remain at ARC's discretion⁹.

In the same way, the ERIH proposed a form aimed at any "Academic, Research and Research-related staff in Universities" and provided details on its site of the procedure to follow for anyone who disagreed with a journal's position in any ratings (ERIH FAQ: "I disagree with the classification of a journal on the ERIH list. What can I do?" and "A journal that I subscribe to/ read is not on the ERIH list. What can I do?"). In addition, the members of the ERIH Steering Committee published a response to their critics in which they encouraged everyone to become actively involved in the revision process:

Despite all warnings, even these initial lists are already being used and will continue to do so. This points to the urgency of improving the available tools, and to the fact that it is in every scholar's interest to contribute to their improvement. [...] ERIH is led by scholars for scholars and the feedback mechanism enables scholars to communicate their views directly to the Expert Panels, and to the Steering Committee. The feedback that is being submitted by individual scholars and scientists as well as by expert communities will have an important role in the updating of the "Initial Lists" (ERIH Reply, 2008: 2 and 3).

Moreover, the ERIH reversed some of its initial stances, particularly in relation to criticisms of journal categories. At first, the members of the Steering Committee maintained that the chosen letters were pertinent for distinguishing between journals publishing "high-ranking international publications" (A), "standard international publications" (B) and those with an "important local/regional significance" (C).

Alternative denominations for the categories had been suggested on repeated occasions – such as: W (world) – I (international) – N (national and regional), or colour-coded identifications – but have been rejected by the Steering Committee on the grounds that ESF Member Organisations had initially approved the A-B-C categorization (ERIH RevSumGuide, 2007: 4).

⁹ An exercise that we undertook in order to monitor the way in which our proposals were taken into consideration. The two French journals which we suggested are finally present, but with a different score.

But confronted with the possible incorrect uses of these categories to assess the quality of academic productions, particularly by ignoring articles published in journals scored as “C”, the ERIH announced that it was seriously considering the idea of replacing scores with written descriptors (Times Higher Education, 2009). The idea would be to conserve the categories in a different form whilst at the same time reminding people that their differences were not the basis for any hierarchy:

Nevertheless, we understand and accept that these three categories could be misperceived as a kind of ranking. The ERIH is therefore in the process of rethinking this categorization in A, B, and C, and of remodelling it according to a division into international, national or regional (ERIH Reply, 2008: 3).

AERES did not practice any form of formal journal consultation, but in October 2008 it conceded the possibility of creating journal lists without any internal hierarchy. Then, during a meeting of thirty or so French journals on the 4th December 2008, AERES’ SSH delegate introduced the positive notion of “perimeter of scientificity”. The idea was to return to a simple list, the parameters of which would be defined by the appropriate scientific communities. Although journals felt that numerous questions remained (inclusion and revision criteria) they reacted positively towards this proposal (Picard, 2008).

From design to content: the revised lists

Revision of the lists was thus organised in accordance with procedures which had not been foreseen by the organisations. Having been improved through its consideration by the journals, it then underwent the development of particular tools such as forms and promises of the clarification or easing of certain criteria. But what was the extent of these revision procedures? How did contributions from the journals and researchers in fact modify the initial ratings?

Although the ERIH showed it had taken criticism into account by making several changes, the new lists have still not appeared. In May 2010, the ERIH spoke about the progress made in its work on ratings and justified the postponement of publication by the fact that it was impossible to fully fund the process. It also declared that the panels had accompanied their rating with a “report from a chair commenting on the revision process” (ERIH Report, 2010: 1).

As for AERES, the revision process came to an end as early as March 2009 for history journal ratings, with the other discipline lists being gradually published throughout the autumn of the same year. These updated lists marked the arrival of new types of relationship between the general framework defined by the Agency and the options selected by each individual discipline. We find three main differences. First, the lists are now presented independently in a PDF file which uses the same layout with the Agency logo and the following title: “List of AERES journals for X field (updated 00.00.2009)”. There then follows a polyphonic argument where AERES’ utterances (identical from one list to another) are in italics, with the contribution of the expert committee inscribed in regular letters. This argument is followed by a list of journals which defines the field’s scientific perimeter.

Second, the arguments accompanying the lists are far more explicit. The AERES text now contains four main points: a reminder of the importance of books which, in numerous SSH, “remains the principle form for scientific works”, the representativeness of the panels whose members come from national research assessment bodies (CNU, CoNRS, IUF) and from qualified foreigners, and the announcement of an annual revision. The contribution of the expert panels is divided into two parts: the first makes public the names of the “revision

commission”, the second provides details of the “modalities for shaping the list” used by the expert panels, which in certain cases specifies the criteria selected for journals to be included in the updated list (e.g. the Sociology-Demography, Philosophy, Anthropology lists).

Finally, the revision process has led to the production of two types of list depending on the discipline-based groups¹⁰. Certain panels have followed AERES’ recommendations by defining a strict scientific perimeter, i.e. without any sort of journal ranking, considering the list “to include all titles to be systematically taken into account when determining the number of publishing academics” during the assessment of research units” (AERES SocioDemoList 2009). Other panels have taken up this principle of a perimeter whilst at the same time adjusting it through internal segmentations, thus integrating a hierarchy of publication outlets. This is for example the case for the History and Philosophy of Science list, where journals are divided into two ranks (1st and 2nd) to provide “a relative indication based on the journal’s quality and influence” (AERES HistPhiloSciList 2009), or for the Psychology-Ethology-Ergonomics list which uses the impact factor to position journals in relation to one another by giving them the “following six descriptors: impact unknown, excellent impact, very good, good, quite good and moderate” (AERES PsyList 2009).

From a quantitative point of view, the lists have undergone two distinct changes: the majority of disciplines have seen a reduction in the number of their journals (e.g. Sociology -7%, Theology -22%, Anthropology -27%). In such cases, the new panels have taken the initial lists and have chosen to include some of the journals previously given a C rating in the new scientific perimeter, with the remainder disappearing. Conversely, for certain disciplines which previously had very short lists, the number of journals taken into consideration has increased dramatically: the Political science list has risen from 25 to 325 journals, that of Educational Sciences from 32 to 530. In the latter case, the panels have considerably internationalised the origins of the journals and included a large number of inter-disciplinary journals.

As far as ARC is concerned, the formal consultation procedure led to a significant increase in the number of SSH journals on the list published in March 2010: from 10,241 journals on the draft lists, to 12,976¹¹. This average increase of one quarter hides very considerable variations between disciplines: whilst most of them saw their number of journals increase (e.g. Psychology +14%, Law +17%, Political Science +30%, Sociology +77%), some – such as Linguistics (-2%) and History (-5%) - saw reductions due to the journals being reassigned to other categories. These quantitative changes are not linked to the journals being divided into tiers. History, for example, fairly stable in terms of the number of journals, significantly modified the scores by moving away from the previously defined quotas: the percentage of journals with a C rating fell from 48% to 38%. On the other hand, the Sociology panel, whose results had not respected ARC’s quotas in 2008, used the increase in the number of journals to draw closer.

But is this formula for division into four tiers still in force? Earlier on we saw that the dual definition, intrinsic and relational, of the ARC categories generated tension. With the revised lists, only the intrinsic definition is maintained. This disappearance of quotas thus eliminates part of the process of list creation, the constraint of division having been very much taken into account (Abelson, 2009; Fisher, Shanks and Lamp, 2009). It would also seem to be allowing certain panels to assert their freedom of judgement, directly based on their expertise.

¹⁰ Qualitative and quantitative elements are based on the comparison of the new lists published between March 2009 and February 2010 with the ones published in July 2008.

¹¹ All of the quantitative elements which follow are based on the use of the John Lamp site (<http://lamp.infosys.deakin.edu.au/era>) which re-sorted ERA’s lists in accordance with FoR.

Discussion: Are journal ratings an alternative policy tool?

The production of journal lists for SSH is a complex process, subject to broad criticism supported by scientific communities who are generally opposed to bibliometric tools. Considering the intensity of these debates, we might wonder what the future of journal ratings will be. Firstly, the cognitive gain from implementing such journal ratings appears to be something that assessment organisations find attractive and will try to extend to other outlets. Secondly, conceived of as a new policy tool, journal lists are exposed to the same criticism as metrics tools and one can expect the limits of their uses to be clearly stated. Thirdly, contrary to IF, which benefited from a long period of monopoly, journal ratings were from the very outset in a situation of competition, of plurality.

The potential extension of ratings

Bibliometrics and scientometrics tools were developed from a journal-based perspective that explicitly evacuates the relevance of other kinds of outlets. Formally produced to classify SSH journals, ratings and lists could be limited in the same way. However, the extension of these tools to other scientific works is not only conceivable, but is even on the agenda of assessment organisations: ARC produced conference ratings at the same time as its journal ratings, and both ESF and AERES wish to do the same with SSH books, either by considering the publishers – as Giménez-Toledo and Román-Román (2009) proposed, or by refining the analysis at collections level. In this way, all forms of academic production in the field of SSH could be quantified or even classified, finally placing these disciplines at the same level as other sciences. Whilst this objective must be taken seriously, it being clearly stated by the organisations and underpinned by scientific evolutionism (Cole, 1983), we must not ignore the backlash produced by these new mechanisms.

Indeed, ARC already uses the same grid for journals, even in natural sciences, recoding the bibliometric data to form four hierarchical categories defined by quotas. This type of transformation of a continuous measurement tool into a discreet rating is now being proposed by bibliometrics companies. Based on methods developed by independent scholars in scientometrics (Magri and Solari, 1996; Solari and Magri, 2000; Wagner, 2009), it has been partially adopted by ISI and by its main competitor, through its own impact measurement, the SCImago Journal & Country Rank (SJR):

In order to offer a framework to facilitate journal assessment, SJR Journal Profile Pages show now journal's SJR indicator Quartiles for each subject category in which they are included. It can be viewed in the small icons alongside Subject Categories Labels stating Q1, Q2, Q3 and Q4 (SJR Indicator Quartiles)¹².

Even if this quota approach is still open to criticism due to the arbitrary basis for its cut-off point¹³, it is a first attempt at universal journal ratings, though on an automated basis of citations and not through peer-reviewing. In particular, contrary to IF-like metrics, it enables an easy comparison between different disciplinary outlets, which is clearly searched by evaluation organisations.

Are journal ratings just bibliometrics?

The use of journal ratings as a policy tool thus appears to be facilitated by the end of the IF monopoly. Nevertheless, the criticism of the use of IF has also been used against them by

¹² <http://www.scimagojr.com/news.php?id=117>.

¹³ See the sociology example where a journal with an SJR of 0.028 is in Q3 and another with an SJR of 0.027 is in Q4.

scientific communities: by taking journal quality as a proxy for the article quality, one carries out an operation of homogenisation through the use of mean. And just as a journal's high IF does not imply that every article will be very much cited and vice versa (Adler et al, 2008), so the presence of an article in a low-rated journal does not say anything about its quality (Singh et al, 2007; Starbuck 2005):

All Panels emphasise that high quality research can appear in journals within all three categories (A/B/C) and that the category itself does not imply a value judgement about the quality of any particular piece of research (ERIH ExPanels, 2008: 4).

Just as the development of direct citation measurements for articles and authors, the h index in particular, makes the publication outlet irrelevant, taking this announcement seriously reduces to nothing a journal rating's usefulness as a tool for quality assessment!

Moreover, the design of the journal lists in no way prevents their quantitative uses, in two different forms. Firstly, as far as AERES and maybe the ERIH are concerned, it is a case of lists which put all works which appear in the selected journals on the same footing. In this sense, journal ratings tend to favour an instantly quantitative approach, as seen in Australia in the 1990s. Concrete uses may then vary: minimum number of articles for those who publish, number of articles per researcher, etc. Secondly, when lists are clearly separated into tiers, as in the first AERES version or in the case of ARC, a scientific hierarchy is preferred, quantitative logic being second to peer review on the quality of the journals. However, as we have shown (Pontille & Torny, 2010), this does not prevent the different actors from reconstructing metrics by introducing equivalence formulae such as $1A = 2B$ or $1A = 3B$

Anticipating these kinds of uses, critics of these tools denounced their role in the production of a hyper-normal science via several mechanisms. For example, they would favour the existence of an annuity for the older journals already included in the rankings (Macdonald and Kam, 2007) and would destroy the younger and supposedly more innovative journals. In constructing the unified representation of a hierarchy, they would favour, like the use of the IF, a race to publication in the high-quality journals, standardising their content: for example, they would force heterodox articles out to the fringes in economics (Lee, 2006; 2008) or in marketing (Svensson, 2006). We have seen that the production of non-ordered lists of journals and of explicit inclusion criteria met these criticisms and appeared to receive approval from the relevant scientific communities.

Towards a multi-rated world

To this criticism of a managerial assessment approach to science and its deleterious effects, can be symmetrically added criticism of the utility of such grids, from the point of view of the political body making the assessment. The most often cited example is the Law discipline: the concrete issues which Australian legal experts can address and which provide the basis for their social utility will not be raised in articles published in British or American journals (Svantesson, 2009). In favouring the latter, ARC's journal ratings would lose sight of their principle mission.

More generally, it is the measurement of the social impact of SSH articles and works which is under discussion, on the understanding that once again the measurements generally used – usually patents – are rarely pertinent. As social actors, and public actors in particular, are not in the slightest bit interested in the publication outlet, being more directly concerned by the content of the articles and its transferability, the local and adapted nature of which is of greater importance, it is other measurements which will need to be developed, such as the “impact criteria” envisioned by the defunct RQF (DEST, 2005) or, more generally, S&T indicators (Grupp and Mogege, 2004). Such a vision thus favours the creation of local ratings

(preferred themes, impact on public policies), be it with the objective of providing financial support for the journals or of allocating resources to research departments, and explains the vast differences between national ratings (Alexander et al, 2007).

Nonetheless, even if their framework of use is very well-defined, as is the case with ARC, once constructed, the management tool breaks away and not only becomes autonomous (Berry, 1983), but is also used in strategic ways (Espeland and Sauder, 2007; Macdonald and Kam, 2007). To the immense surprise of their promoters, the Australian rankings are of interest to the rest of the world! Similarly, the rankings of economics and management journals, produced by German, British, French, Dutch, etc. organisations, were aggregated in the Journal Quality List, produced by Anne-Wil Harzing, a Professor in International Management at the University of Melbourne, whose approach was motivated by her former organisation's poor ranking of journals in her field. So, while criticism remained important in our three cases, at least some sections of scientific communities (publishers, editors, learned societies, scholars) have already taken into account the multitude of ratings. Though they could be subsumed to build integrated balanced ratings (Moed, 2007), their diversity is also a powerful resource. As much as economic actors have created a profusion of metrics (Van Noorden, 2010), the many political bodies implied in research assessment have created an emergent multi-rated world.

Conclusion

The production of journal ratings is far from being a mere inventory-making exercise – indeed, it is an eminently cognitive and political task. It involves just as much the determination of pertinent categories with which to define the internal division of elements and the introduction of specific inclusion and exclusion criteria, as the more or less explicit definition of the permitted and proscribed uses of ratings. And as we have seen, the modalities selected by ARC, ESF and AERES produced significantly different effects: the boundaries of discipline-based groups, the modalities for consulting scientific communities, the revision processes adopted to respond to the high level of criticism, and the very composition of the journal lists have led to significant variations. Despite this, the common objective is to develop a policy tool which is suited to SSH and which is intermediate between peer review and sundry metrics.

As an alternative tool, journal lists and ratings are subjected to a dual tension. Firstly, as something between a political instrument and a scientific assessment, they are the object of gradual compromises between scientific communities and specific research management organisations. Once such a compromise has stabilised, at least temporarily, its material form is subject to a second tension between predefined local uses and a universal tool, as each journal, each community, each panel and each organisation can use it as a black box to count or to qualify scientific works. As the number of journal lists and ratings and their uses grow, these intrinsic tensions will not only feed the need for revisions in search of the perfect assessment tool, but also fuel critique and protest which remain at the heart of many SSH.

Acknowledgements

We are very grateful to Chris Hinton for his translating assistance.

Appendix. Data sources (all url accessed 22 April 2010)

Australian Research Council

- Abelson P., 2009: “The Ranking of Economics Journals by the Economic Society of Australia”, *Economic Papers*, vol.28 (2), p. 176–180.
- ARC Tiers, 2008: “Tiers for the Australian Ranking of Journals” (http://www.arc.gov.au/era/tiers_ranking.htm).
- DEST, 2005 : “ARC Support Paper: Assessing research impact in the RQF”, September 2005, 10 pages (<http://www.dest.gov.au/NR/rdonlyres/F5166FC3-B20C-4931-AD6E-B86AAFEF96F6/7862/Assessingresearchimpact.pdf>).
- DEST, 2006: “Quality metrics”, September 2006, 7 pages (http://www.dest.gov.au/NR/rdonlyres/EC11695D-B59D-4879-A84D-87004AA22FD2/14099/rqf_quality_metrics.pdf).
- Donovan, 2005b : Donovan C., “A Review of Current Australian and International Practice in Measuring the Quality and Impact of Publicly Funded Research in the Humanities, Arts and Social Sciences”, REPP discussion paper 05/3, November 2005, 37 pages (http://repp.anu.edu.au/papers/200511_disc03_review.pdf).
- EPAT, 2008: “Editorial: ERA Journal Ranking Exercise An Open Letter to the Australian Research Council”, *Educational Philosophy and Theory*, vol.40 (7).
- Fisher J., Shanks G., Lamp J., 2008: “ERA replaces RQF”, *Australasian Journal of Information systems*, vol.25 (2), p. 5-6.
- Hainge G., 2008 : “Society for French Studies slams ERA journal rankings”, *The Australian*, 12th August 2008.

European Science Foundation

- ERIH ExPanels, 2008: “Expert Panels: process and methodology of selection”, 8 pages.
- ERIH FAQ, 2008: Frequently Asked Question (<http://www.esf.org/research-areas/humanities/erih-european-reference-index-for-the-humanities/frequently-asked-questions.html>).
- ERIH SumGuide 2007: “Summary guidelines”, September 2007, 2 pages (http://www.esf.org/index.php?eID=tx_nawsecuredl&u=0&file=fileadmin/be_user/research_areas/HUM/Documents/ERIH/ERIH%20summary_guidelines_Sept_07.pdf&t=1273323337&hash=0507fa86a5cb6038fa52192539a0e959).
- ERIH RevSumGuide 2007: “Revised ERIH summary guidelines”, October 2007, p. 4 (http://www.esf.org/index.php?eID=tx_nawsecuredl&u=0&file=fileadmin/be_user/research_areas/HUM/Documents/ERIH/Scope%20Notes/ERIHRevised_GUIDELINES.pdf&t=1271954354&hash=d4294d8851b93806448537b59ef7c06e).
- ERIH Reply 2008: “A reply to the Criticism”, December 2008, 3 pages (<http://www.esf.org/research-areas/humanities/erih-european-reference-index-for-the-humanities/joint-response-to-criticism.html>).
- ERIH Report 2010: “Report 2008-2009”, 32 pages (<http://www.esf.org/research-areas/humanities/erih-european-reference-index-for-the-humanities.html>, accessed 6 May 2010).
- Kritika 2009: “Failing the Grade. The Craze for Ranking Humanities Journals”, *Kritika: Explorations in Russian and Eurasian History*, vol.10 (1), p.1–6.
- Times Higher Education 2008a: “Historians decry journal rankings”, 4 January.
- Times Higher Education 2008b: “Philosophers in ranking protest”, 28 August.
- Times Higher Education 2009: “Index of journals scraps controversial grades”, 22 January.

The French Agency for Evaluation Research and Higher Education

- AERES GenList 2008: Argument accompanying the full journals list, 28.07.2008 (<http://www-aeres-evaluation-fr.aw.atosorigin.com/La-liste-des-revues-scientifiques>).
- AERES SciPoList 2008: Argument accompanying the Political Science journals list, 29.12.2008 (<http://www-aeres-evaluation-fr.aw.atosorigin.com/Science-politique>).
- AERES Petition 2008: “Pour le retrait complet et définitif de la ‘liste des revues’ de l’AERES”, October 2008 (<http://www.appelrevues.org/>).
- AERES HistPhiloSciList 2009: Argument accompanying the History and Philosophy of science journals list (http://www.aeres-evaluation.fr/content/download/12884/180879/file/20091015095316_-_histoire_pour_mel.pdf).
- AERES PsyList 2009: Argument accompanying the Psychology-Ethology-Ergonomics journals list (http://www.aeres-evaluation.fr/content/download/12886/180893/file/091014_ListePsycho.pdf).
- AERES SocioDemoList 2009: Argument accompanying the Sociology-Demography journals list (http://www.aeres-evaluation.fr/content/download/12888/180907/file/Liste_revues_socio_demo_dec09.pdf).
- Annales HSS 2008: “Editorial: Classer, évaluer”, n°6, p. 1-4.
- Briatte F., 2008, “Comparaison inter-classements de revue en sociologie-démographie et en science politique”, *Bulletin de méthodologie sociologique*, n°100, p. 51-60.
- Merllié 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*, Tome 134 - n°1, p. 91-99.
- Norois 2008: “Editorial”, n°208, p. 7-9.
- Picard E., 2008, “Compte-rendu de la réunion des revues du 4 décembre 2008” (<http://www.afsp.msh-paris.fr/observatoire/etudes/evaluation/reunionrevues041208.pdf>).
- RHMC 2008 : Revue d’Histoire Moderne et Contemporaine, “La fièvre de l’évaluation”, n°4-bis, 2008.

References

- Adler, R., Ewing, J., Taylor, P., 2008. Citation statistics. Rep. 41, International Mathematical Union.
- Alexander, J.K., Lecoutre, M., Scherer, R.F., 2007. A Global Comparison of Business Journal Ranking Systems. *Journal of Education for Business*, **82**(6), 321-327.
- Auranen, O., Nieminen, M. 2010. University research funding and publication performance – An international comparison. *Research Policy*, **39**(6), 822-834.
- Bazerman, C. 1988. *Shaping written knowledge: the genre and activity of the experimental article in science*. Madison, Wisconsin: The University of Wisconsin Press.
- Berry, M., 1983. Une technologie invisible ? L’impact des instruments de gestion sur l’évolution des systèmes humains. Paris, <http://crg.polytechnique.fr/fichiers/crg/publications/pdf/2007-04-05-1133.pdf>
- Bollen, J., Van de Sompel, H., Hagberg, A., Chute, R., 2009. A principal component analysis of 39 scientific impact measures. *PloS one*, **4**(6), e6022.
- Bowker, G.C., Star, S.L., 1999. *Sorting things out. Classification and its consequences*. Cambridge, MA: MIT Press.
- Brinn, T., Jones, M.J., Pendlebury, M., 2000. Measuring research quality: peer review 1, citation indices 0. *Omega*, **28**(2), 237-239.
- Butler, L., 2004. What happens when funding is linked to publication counts? In *Handbook of Quantitative Science and Technology Research*, eds. H.F. Moed, W. Glänzel, and U. Schmoch, pp. 389-405. Kluwer Academic Publishers, Dordrecht.

- Butler, L., 2008. Using a balanced approach to bibliometrics: quantitative performance measures in the Australian Research Quality Framework. *Ethics in Science and Environmental Politics*, **8**, 83-92.
- Butler, L., McAllister, I., 2009. Metrics or Peer Review? Evaluating the 2001 UK Research Assessment Exercise in Political Science. *Political Studies Review*, **7**(1), 3-17.
- Campanario, J.M., 1998. Peer Review for Journals as it Stands Today--Part 1. *Science Communication*, **19**(3), 181-211.
- Castles, I., 1993. Australian standard research classification. A set of classifications for R1D projects. Australian Bureau of Statistics, catalogue n°1297.0.
- Charlton, B.G., Andras, P., 2007. Evaluating universities using simple scientometric research-output metrics: total citation counts per university for a retrospective seven-year rolling sample. *Science and Public Policy*, **34**(8), 555-563.
- Cole, S., 1983. The hierarchy of the sciences ? *American Journal of Sociology*, **89**(1), 111-139.
- Donovan, C., 2005a. The governance of social science and everyday epistemology. *Public Administration*, **83**(3), 597-615.
- Espeland, W.N., Sauder, M., 2007. Rankings and Reactivity : How Public Measures Recreate Social Worlds. *American Journal of Sociology*, **113**(1), 1-40.
- Garfield, E., 1970. What is a Significant Journal? *Current Contents*, **18**, 106-107.
- Giménez-Toledo, E., Román-Román, A., Alcain-Partearroyo, M.D., 2007. From Experimentation to Coordination in the Evaluation of Spanish Scientific Journals in the Humanities and Social Sciences. *Research Evaluation*, **16**(2), 137-148.
- Giménez-Toledo, E., Román-Román, A., 2009. Assessment of humanities and social sciences monographs through their publishers: a review and a study towards a model of evaluation. *Research Evaluation*, **18**(3), 201-213.
- Gläser, J., 2004. Why are the Most Influential Books in Australian Sociology not Necessarily the Most Highly Cited Ones? *Journal of Sociology*, **40**(3), 261-282.
- Gläser, J., Laudel, G., 2005. Advantages and dangers of "remote" peer evaluation. *Research Evaluation*, **11**(2), 141-54.
- Grupp, H., Mogege, M.E., 2004. Indicators for national science and technology policy. Their Development, Use, and Possible Misuse. In *Handbook of Quantitative Science and Technology Research*, eds. H.F. Moed, W. Glänzel, and U. Schmoch, pp. 75-94. Kluwer Academic Publishers, Dordrecht.
- Hix, S., 2004. A Global Ranking of Political Science Departments. *Political Studies*, **2**, 293-313.
- Hogler, R., Gross, M.A., 2009. Journal Rankings and Academic Research: Two Discourses About the Quality of Faculty Work. *Management Communication Quarterly*, **23**(1), 107-126.
- Kalaitzidakis, P., Mamuneas, T.P., Stengos, T., 1999. European economics: An analysis based on publications in the core journals. *European Economic Review*, **43**, 1150-1168.
- Laband, D.N., 1985. A ranking of the top Canadian economics departments by research productivity of graduates. *Canadian Journal of Economics*, **18**(4), 904-907.
- Lamont, M., 2009. *How professors think : inside the curious world of academic judgment*. Harvard University Press, Cambridge, Mass. & London.
- Lee, F.S., 2006. The Ranking Game, Class, and Scholarship in American Mainstream Economics. *Journal of Economics*, **3**, 1-41.
- Lee, F.S., 2008. A Case for Ranking Heterodox Journals and Departments. *On the Horizon*, **16**(4), 241-251.
- Leydesdorff, L., 2009. How are New Citation-Based Journal Indicators Adding to the Bibliometric Toolbox ? *Journal of the American Society for Information Science*, **60**(7), 1327-1336.
- Macdonald, S., Kam, J., 2007. Aardvark et al.: Quality Journals and Gamesmanship in Management Studies. *Journal of Information Science*, **33**(6), 702-717.
- Maier, G., 2006. Impact Factors and Peer Judgment: The Case of Regional Science Journals. *Scientometrics*, **69**(3), 651-667.

- Magri, M-H and Solari A., 1996. The SCI journal citation reports: a potential tool for studying journals ? I . Description of the JCR journal population based on the number of citations received, number of source items, impact factor, immediacy index and cited half-life. *Scientometrics*, **35**(1), 93-117.
- Martin, B.R, Irvine, J., 1983. Assessing Basic Research. Some Partial Indicators of Scientific Progress in Radio Astronomy. *Research Policy*, **12**(2), 61-90.
- Moed, H.F. 2007. The Future of Research Evaluation Rests with an Intelligent Combination of Advanced Metrics and Transparent Peer Review. *Research Evaluation*, **34**(8), 575-583.
- Moed, H.F., Luwel, M., Nederhof, A.J. 2002. Towards Research Performance in the Humanities. *Library Trends*, **50**(3), 498-520.
- Moed, H.F., van Leeuwen, T.N., 1996. Impact Factors Can Mislead. *Nature*, **381**(6579),186.
- Moussa, S., Touzani, M., 2010. Ranking Marketing Journals Using the Google Scholar-based hg-Index. *Journal of Informetrics*, **4**(1), 107-117.
- Nederhof, A.J., 2006. Bibliometric Monitoring of Research Performance in the Social Sciences and the Humanities: A Review. *Scientometrics*, **66**(1), 81–100.
- Nisonger, T., 2002. The Relationship Between International Editorial Board Composition and Citation Measures in Political Science, Business, and Genetics Journals. *Scientometrics*, **54**(2), 257-268.
- Norris, M., Oppenheim, C., 2007. Comparing Alternatives to the Web of Science for Coverage of the Social Sciences Literature. *Journal of Informetrics*, **1**(2), 161-169.
- Pons-Novell, J., 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*, **42**(6), 689-699.
- Pontille, D., Torny, D., 2010. Revues qui comptent, revues qu'on compte : produire des classements en économie et gestion. *Revue de la Régulation*, **8**, <http://regulation.revues.org/index8881.html>
- Power, M.P., 1997. *The Audit Society : Rituals of Verification*. Oxford Univeristy Press, Oxford.
- Seglen, P.O., 1997. Why the Impact Factor of Journals Should not be Used for Evaluating Research. *British Medical Journal*, **314**(7079), 498-502.
- Singh, G., Haddad, K.M., Chow, C.W., 2007. Are Articles in "Top" Management Journals Necessarily of Higher Quality? *Journal of Management Inquiry*, **16**(4), 319-331.
- Solari, A. and Magri, M-H, 2000. A new approach to the SCI journal citation reports, a system for evaluating scientific journals. *Scientometrics*, **47**(3), 605-625.
- Starbuck, W.H., 2005. How Much Better Are the Most-Prestigious Journals? The Statistics of Academic Publication. *Organization Science*, **16**(2), 180-200.
- Svantesson, D.J.B., 2009. International Ranking of Law Journals – Can it Be Done and at What Cost? *Legal Studies*, **29**(4), 678-691.
- Svensson, G., 2006. The Paradoxnoia of Top Journal(s) in Marketing. *European Journal of Marketing*, **40**(11/12), 1153-1168.
- Thomas, P., Watkins, D., 1998. Institutional Research Rankings via Bibliometric Analysis and Direct Peer Review: A Comparative Case Study with Policy Implications. *Scientometrics*, **41**(3), 335-355.
- Van Noorden, R., 2010. A profusion of measures. *Nature*, **465**, 864-866.
- White, H.D., Boell, S.K., Yu, H., Davis, M., Wilson, C., Cole F.T.H., 2009. Libcitations : A Measure for Comparative Assessment of Book Publications in the Humanities and Social Sciences. *Journal of the American Society for Information Science*, **60**(6), 1083-1096.
- Wagner, A.B., 2009. Percentile-Based Journal Impact Factors: A Neglected Collection Development Metric. *Issues in Science and Technology Librarianship*, **57**. <http://www.istl.org/09-spring/refereed1.html>
- Wouters, P., 1999. *The Citation Culture*. PhD. Thesis, University of Amsterdam, Amsterdam.
- Zuckerman, H.A., Merton, R.K., 1971. Patterns of Evaluation in Science: Institutionalisation, Structure and Functions of the Referee System. *Minerva*, **9**(1), 66-100.