Giménez-Toledo, Elea (2015). Evaluation of scientific production: short critical analysis. *RELIEVE*, *21* (1), art. M2. DOI: 10.7203/relieve.21.1.5160

Revista ELectrónica de Investigación y EValuación Educativa

RELIEVE

ISSN: 1134-4032

e-Journal of Educational Research, Assessment and Evaluation

Evaluation of scientific production: short critical analysis

La evaluación de la producción científica: breve análisis crítico

Giménez-Toledo, Elea
Consejo Superior de Investigaciones Científicas (CSIC)

Resumen

Se realiza una revisión histórica (centrada fundamentalmente en España, pero con múltiples referencias a otros ámbitos internacionales) del proceso de evaluación de las producción científica, especialmente de las publicaciones periódicas. Se analizan también las alternativas disponibles en la actualidad, tras una etapa convulsa que ha visto desaparecer diversos instrumentos para la evaluación de revistas. Por último se identifican algunas tendencias y necesidades en el ámbito de la evaluación de publicaciones científicas que presumiblemente marcaran el desarrollo futuro de futuros instrumentos y procedimientos.

Palabras clave:

Evaluación; Comunicación Científica; Publicación Científica Revistas; Revistas Científicas

Fecha de recepción
30 de marzo de 2015

Fecha de aprobación
21 de abril de 2015

Fecha de publicación
21 de abril de 2015

Abstract

A historic review (fundamentally centered in Spain, but with multiple references to other international fields) of the evaluation process of scientific production has been carried out, mainly of periodicals. Available alternatives have also been analyzed after a turbulent period that has seen varied instruments for journal revision disappear. Finally, some of the tendencies and necessities in the field of scientific publication evaluation that will presumably mark the future development of future instruments and procedures are identified.

Keywords:

Evaluation Scientific; Communication; Scientific Publishing Journals; Scientific Journals

Reception Date
2015 March 30

Approval Date
2015 April 21

Publication Date:
2015 April 21

Every anniversary is associated with a necessary balance. The commemoration of the twenty years of the journal RELIEVE represents an extraordinary occasion to look back and critically review the scenario of scientific evaluation that so affects the edition of academic journals. The long path of RELIEVE has allowed it to attend to all of the changes that it has produced in the academic field, from the intensification of the scientific evaluations to the appearance of ANECA, from the scientific edition more or less amateur to one of the most professionalized, from restricted access to open access to scientific content, from the existence of a single source of bibliometric indicators to a generation of sources with indicators for national scientific journals.

The scientific publications constitute the fundamental core of the evaluation of the research activity. It is not a new fact but it is necessary to contextualize the object of this article: how the evaluation of publications is dealt with, what indicators are chosen, who produces them, and how it affects all researchers, evaluators, and editors.

Evaluation model of the Spanish evaluation agencies

The creation of the CNEAI (1989); the birth of ANECA (2002); and the imminent merging of both [1], foreseen for 2015; have marked to a certain extent the behavior and the evolution

Corresponding author / Autor de contacto
Elea Giménez-Toledo. Centro de Ciencias Humanas y Sociales (CCHS). Consejo Superior de Investigaciones Científicas (CSIC). Albasanz, 26-28. 28037 Madrid (España). *Correo electrónico:* elea.gimenez@cchs.csic.es

of Spanish scientific journals, especially those dedicated to the fields of the humanities and social sciences. The establishment of criteria for the evaluation of scientific production- a fundamental pillar in evaluation- of professors and researchers is able to translate, in practice, as the establishment of guidelines of edition for academic journals. Only those that fulfill the parameters established by CNEAI and ANECA have the possibility of being well evaluated and, therefore, will be the most demanded by Spanish authors. At the end of the day, the "channel" that they choose (journal, publishing house, etc.) influences largely on their success in the evaluation to which they are subjected.

At the same time, the criteria defined by those evaluation agencies have been based on national science policy and the priorities that they establish, the international practices of evaluation, and the research conducted by different Spanish research groups. Concerning this last point, there is very clear evidence. The information systems and indicators about Spanish scientific journals explicitly mentioned as much by CNEAI as ANECA were created by two research groups: the group EC3 of the University of Granada and the group EPUC [2] of the CSIC. Also, the criteria that forms part of the FECYT quality label, and those that define the scientific journals that CNEAI considers among the "eligible", proceed from the indicators defined by both research groups.

It is important to emphasize this fact for several reasons: evaluation criteria is supported in evidence that results from research; the agencies show their permeability to the research results generated by specialized groups and, therefore, takes into account the scientific/technical criteria in addition to the political; this is the effective transference of the conducted research from the groups to the agencies. The formulation or the application of the criteria is not always as correct or precise as it could be and this at times generates a lack of acceptance on the part of the researchers under evaluation. Nevertheless, it is important to recognize that in most of cases the criteria utilized by the evaluation agency are not improvised and is proposed or supported by the researchers that are part of the thematic panels.

The revision of the different resolutions of CNEAI, published periodically in the BOE throughout the years, allows for observation up to what point the evaluation criteria has evolved, in both transparency and precision as well as in the adaptation of the scientific communication habits of the different fields. Since the initial selected criteria- published or not in journals of the Web of Science- to the criteria distinguished by areas and that consider other sources of information, there has been a clear evolution. That does not mean that the evaluation models are perfect, or that the different scientific communities fully accept the criteria, but rather that there is an interaction, more or less direct, between the experts panels and the researchers (through scientific associations, for example) and with research groups specialized in research evaluation. That interaction allows for evolution in the evaluation criteria and, in some way, better responses to the reality of scientific communication practices of researchers.

Main features of the current system of research evaluation

ANECA and CNEAI, the two scientific evaluation agencies of state level that perform evaluation of researchers (at an individual level), among others, share some very relevant characteristics in the way of evaluating scholarly publications. Analyzing the two reference documents with regard to evaluation criteria, it is possible to identify some common features that, by the way, serve as a reminder to avoid the frequent misunderstandings in research evaluation. In a schematic form, it would be possible to say that both:

- Follow a process of "informed peer review", through which the scientific production of the researchers is evaluated from the judgement of the

experts who make up each thematic panel and by bibliometric indicators that serve as reference or support.

- Have defined distinct evaluation criteria depending on the disciplinary field and, for example, the criteria employed to evaluate Biomedicine are not the same as those used to evaluate Philology. The different thematic committees are integrated by specialists in the different disciplines and they have autonomy regarding the criteria that are applied [3]; it is tos ay that criteria are adapted, at least in part, to the characteristics of scientific communication in those disciplines.
- The channel -not the article, the chapter, nor concrete books- is evaluated. Although there is the possibility of analyzing the content in a certain way (to which they contribute the "defense" that authors of the articles make, in the framework of the evaluation processes of CNEAI, for example).
- In general terms, the journal articles have more weight than books or chapters of books. In the majority of disciplines, it is hoped that researchers generate articles as a result of their research. Only in some disciplines of humanities and social sciences, books or chapters of books are clearly foreseen as research results. The reference texts about the evaluation procedure of the agencies allow inferring the equivalency between the weight of an article and the weight of a book: a book would be equivalent to two articles or an article and a chapter of a book.
- In the evaluation of the books, the different indicators of quality are taken into account but there are just few sources that collect values for these indicators.
- In journal evaluation, the Web of Science and their Journal Citation Reports are the preferred reference sources in the evaluation process. The absence or presence of a journal in these databases that integrate the WoS, as well as the factor of the reflected impact in the JCR, are key indicators- and prevail above the rest- in the evaluation processes of Spain.
- When a publication does not have indicators in WoS/JCR, then other national and international sources are considered. For the humanities and social sciences, ERIH and Scopus are the broader sources in the international framework. ERIH is in the process of change as will be discussed later on, while Scopus continues increasing the number of indexed journals, being thus less selective than WoS and generating some indicators (through Scimago Journal Rank), still not considered expressly by evaluation agencies. In relation with the sources providing indicators for Spanish journals, although referred to and cited more expressly, it is true that many of them have ceased to be operational and up-to-date, because of a lack of funding. This matter will be also discussed later.
- Although the law of science represents a clear bet for open access, the scientific evaluation is based on "closed" sources. In the data that Delgado (2015) offers, only 10.6% of the journals collected in JCR are open access. The most valued journals (according to JCR) are not OA or are hybrids. This means that the research must choose

  - For full open access, that is to say, choosing a journal in open that doesn't charge for publishing. This way, it guarantees the free availability of the research results, as the Law of Science promotes; nevertheless, this action could have negative influence in evaluation processes.
  - For the payment of open access publication, removing resources from

the research funds. In this way, resources will be devoted to fund large publishing groups and not research.

- For publishing in a closed journal (which is accessed by subscription), usually belonging to large publishing groups and collected in JCR, which facilitates a positive evaluation.

A critical review of the sources on journals

There is little to say now about the main source of information employed for evaluation (WoS). There is a huge amount of studies based on scientific production, citations and impact from the WoS. There also too many others that demonstrate the limitations of this source for evaluation of some areas and the precautions that they should take. Likewise the characteristics of Scopus are well known. Although it is a secondary source according to the hierarchy of sources that establish the evaluation agencies in Spain, the largest range of journals and the indicators that it provides directly or indirectly, make it a common source in bibliometric studies.

There is no doubt that the internationality and the multidisciplinarity of both make then well suited to make productivity and impact analysis of all kinds. Neither is there any doubt of the interest that governments have, since they permit then to easily obtain indicators that are comparable among countries, and therefore, have standards and objectives that they aim for.

Nevertheless, although this point of view is understandable, there is also the fact that they don't well represent the totality of scientific production of quality that is generated in different disciplines and in different countries. For that reason, many countries of the world and some "regions" (Europe with ERIH or Latin American with Latindex) proposed and continue to propose the creation of alternative, national, or supranational systems of indicators for the scientific production generated in each country, at the edge of the collection in WoS or Scopus. These initiatives are able to be summarized in the following models:

- Development of CRIS (Current Research Information Systems) at a national level that, in addition, entail certain quality indicators for scientific production collection. This is the case of the Belgian or Norwegian system (Siversten, 2010).
- Categorizations of journals that are applied to scientific production. This was the case of the ANEP in Spain, and is the case of the ANVUR in Italy, AERES in France, and Publindex in Colombia, to give just a few examples. ERIH could also be included in this paragraph, although taking into account the reserves that their creators have always had regarding the use of the categorization for individual level individual evaluation.
- Evaluation systems of scientific journals, integrating different indirect qualities; among which the regional level, Latindex, can be found, which encompasses the journals of Latin America, the Caribbean, Spain, and Portugal, and those of national level, RESH and DICE.
- Evaluation systems that show a selection of journals that meet the defined criteria, although they do not have the role of showing the value of indicators (the compliance of the same) for each journal[4]. This is the case of the so-called "core" of journals created in some Latin American countries or could even be considered here the journal sources included in Scielo or Redalyc.
- National impact indexes: Polish Sociology Citation Index, Russian Citation Index or the Spanish In-RECS, In- RECJ and In-RECH

These examples are sufficient to show that the evaluation of national scientific journals is not a fancyof a country in particular, but rather one that responds to a clear need for having indicators that provide information about the publishing sector of the journals of a country, which will help to make decisions in scientific

evaluation, reporting on what is not covered by WoS and that, in addition, serves for other purposes such as fixing certain quality standards or promoting improvements in the publication. Without national systems of evaluation, results will be biased, and on the other hand, the knowledge of the publishing sector will be deficient.

Latin America was among the pioneers in the definition of quality criteria for scientific publications and national evaluation systems that have been developing in an unequal manner depending on the time and the country. In Spain, the work on this question began to consolidate in the late 90's. Between 2004 and 2006 the first versions of RESH, DICE, and In Recs had already been made public. They would then begin to appear other systems on on journals quality, , such as the still existing MIAR, and to create initiatives from the most political/institutional fields, like the FECYT label of quality. It is important to emphasize the origins, because in that time there were few European countries that had their own assessment tools, although they already were beginning to take shape.

Knowing this evolution, it does seem paradoxical that a decade later none of the pioneer systems in Spain and Europe (RESH, DICE, and In Recs) are already operative and that the reason is not scientific but rather economic: a lack of the funding necessary to maintain them. Meanwhile, different European countries have already created their own systems that have important institutional support and that guarantee, in a way, their viability. This fact offers an interesting standard for the future. Although the research groups have capabilities and competencies for developing systems of publication evaluations, public institutions -and specifically those that have responsibilities in the field of scientific policy- should promote, sustain, and support the systems. After all, the information that these systems offer is a public service that serves many agents in the academic field, and on the other hand, play the role of "protecting" the scientific heritage built by journals, or give the opportunity to high quality journals not indexed in the WoS the to be well evaluated. It should not be forgotten that universities budgets support many institutional journals that, paradoxically, are undervalued in some evaluation processes. Here occurs an inconsistency between the achieved efforts for editing, by means of the investment of public resources, and the recognition in terms of evaluation; this especially affects the highly specialized journals (Mañana-Rodríguez, 2013), both geographically and thematically.

In defense of those systems it should be said that they were systems considered as reference sources in Spanish evaluation agencies; in fact, the still are, despite the fact that their information is out of date. This fact allows, for the first time, evaluation of a set of scientific production much wider than that which is included in the Web of Science, a question that is fundamental in humanities and social sciences. The report of CRUE (Michavila, 2012) serves as a forceful example. According to data offered- referring to the scientific production generated by the university community and obtained in 2010- the publication of articles in art and humanities involved 44% of the total scientific production of the area (7763 articles.) Amongst them, only 15% are articles covered by the WoS. This means that not using other evaluation sources complimentary to the WoS is denying 85% of the articles produced in this discipline a chance for evaluation (or they are evaluated without tools that guarantee certain objectivity.) Perhaps this quantitative data relating to part of its works that stay invisible or undervalued without these evaluation tools is expressive enough.

It is important to add that the evaluation systems of journals do not serve just for evaluation purposes. They also serve to have a more comprehensive knowledge of the publishing sector of a country, to better understand the behavior of the disciplines, to identify the rigor of the publication and/or the professionalism, etc. All of these are questions that would be or should be of interest to those

who are allocating resources to the publication of journals; to the publishers themselves, that are able to "see themselves" in relation to other publishers, and that, in addition, have a frame of reference of the international standards of scholarly publishing, a horizon towards which it approaches; to the authors that will have the best knowledge of the journals in their area; and, of course, to the science policy-makers that, in addition to counting on tools for scientific evaluation, would be able to design support programs for the publications.

In Spain, the case is that these systems have disappeared (or they do not have the utility that they once had) and that, in a way, a regression has been produced. Although the reference documents of the agencies continue demonstrating them as tools complimentary to evaluation, the fact is that its irrelevancy causes the evaluation of scientific production to return to being based almost exclusively on the WoS and Scopus as well.

It should be said that neither effort nor time is lost in these projects. The publication of scholarly journals has taken an important qualitative leap. In the first place, because they publish much better than before; from the formal point of view, most of journals follow international guidelines of scientific publishing, something that didn't happen fifteen or twenty years ago, amongst other things because they were unknown; from the point of view of content quality things have also advanced because many of the journals have substantially improved their political publishing and their filters. There should be no doubt that there is a greater and better knowledge of what is involved in good scientific publishing. Another different thing is that all the journals that are published follow the best paths and concern themselves with maximum quality of the articles. In the same way that some publishers have seen an opportunity to review and think over their publishing projects in the establishment of indicators and in scientific evaluation, advancing towards better quality content and towards the professionalization of the publication, others have opted only to feign the observance of indicators but without modifying the editorial practice; this option is so visible for those who work with indicators such as for the evaluators.

Current tools complimentary to WoS and Scopus

Still two systems remain that, being demanding in a manner comparable to WoS, could fill the space left by the evaluation systems of national journals. One of those is the European ERIH. The European Reference Index for the Humanities that promoted and developed the European Science Foundation is now already a product transferred to the Norwegian Social Sciences Data Services. It has changed producers and also the denomination ERIH Plus is now a system that includes journals of the humanities but also social sciences. In the moment of writing this article, the information offered for each journal is practically the editorial and, in some cases, if it is a "peer reviewed" publication. Of course the system is in transition because new journal applications are evaluated for entering the system and, on the other hand, must hope that they are offered some qualitative information about the publications or rather that the process of selection itself is so rigorous that the only presence of the journal in ERIH Plus is an unequivocal sign of quality. The absence or presence of a journal in ERIH Plus would be the only way to utilize, currently, the index of evaluation effects. Nevertheless, despite the fact that the system is able to be one of the clear guides for the evaluation of the outputs in humanities and social sciences, it is in the transitory phase and still needs to consolidate itself.

The other system is CIRC, a journal categorization that results from the addition of contained information in different sources of information about journals. It is like a summary of what other databases say of a journal. It was the result of a scientific collaboration amongst various Spanish research groups (Torres-Salinas et al., 2010)

although afterwards it became a development of the company EC3 metrics. National and international sources were integrated in its origin, which made it an efficient and solid source for evaluating scientific production of a department, a research institution, etc. Presently, the producers have announced an imminent publication of a new version. The predictable changes that are to come are clear: without national sources of indicators, CIRC would be based more in the existing international sources that, without losing interest in the tool, represent again a step backwards, because the evaluation of a large number of Spanish journals are excluded.

Another existing source of data utilized more and more, although not officially, is Google Scholar. It does not have the "closed" character of the rest of the sources because it constantly updates its data and information and does not select journals, so that it provides information for all. From Google Scholar data can be obtained from journals and as well as rankings, as the Statistical Cybermetrics Research Group of the University of Wolverhampton, the group EC3 of the University of Granada, and the application Publish or Perish are doing (Harzing, 2007). Without a doubt, it is an additional source for scientific evaluation that represents a large advantage: the production of the "databases" depends on a large company, and therefore is a matter of exploiting the data. On the other hand, it also has related inconveniences with the value or the representation of the quotes in the humanities and the social sciences and the possibility of manipulating the information contained in Google Scholar (Delgado *et al.*, 2014).

In whatever case, the ease with which the impact indicators are obtained and its availability consolidates its use in some countries.

In conclusion

Analyzing the evaluation of scientific journals that have been in Spain since the evaluation agencies increased their activity- in the case of CNEAI- or created themselves- in the case of ANECA-some conclusions can be drawn.

In the first place, although it is obvious to say it, it is clear that the Web of Science, and its JCR, are going to continue as reference sources in evaluation. Having become an international reference source, its use allows the comparability of evaluation processes and the requirements to researchers; it is a standard that many countries are willing to share. If this is coupled with the tradition or inertia in its use, the settlement of the tool, the possibility of modulating the demands for areas depending on the quartiles, the leadership of Thomson Reuters and its influence in evaluation policies, it appears impossible to envision a scenario in which the WoS is not the key reference source in evaluation processes, at least in some countries, like Spain, evaluations process based on the denominated "informed peer review" are followed, that is, those that combine the use of bibliometric indicators with the most qualitative trial of a panel of experts. A model exists in the United Kingdom that in the defined criteria for its Research Excellence Framework (REF), expressly excludes the use of bibliometric indicators, data about quotes and journal categorization for the evaluation of scientific activity that occurs. The evaluation system is qualitative: it is based in the trial of the specialists in the area. Humanists and social scientists have been claiming qualitative assessments opposite to the quantitative ones and it seems that a system such as the British could enjoy greater acceptance. But also, it is necessary to indicate that these systems are very expensive and that, if anything, they will only be allowed in those countries with an intense activity and investment in research.

The middle way is constituted, undoubtedly, by the categorizations and evaluation systems of journals that cater to different publication characteristics- qualitative and quantitative- while at the same time represent an opportunity to know the publishing industry as a whole, and

consequently, the deficits that may exist and the actions that can be carried out to defend the good scientific publication which has no place in international databases.

References

Cetto, A. M., Alonso-Gamboa, J. O., Córdoba González, S., Giménez-Toledo, E. & Chávez Sánchez, G. (2012). Organized access to the Ibero-American quality journals: The PPL experience. *Scholarly and Research Communication, 4* (1), p.19.

Delgado López-Cózar, Emilio (2015). Open Access E-journals: Past, Present, and Future. *RELIEVE*, *21* (1), art. M1. DOI: 10.7203/relieve.21.1.5005

Delgado López-Cózar, E., Robinson-García, N. & Torres-Salinas, D. (2014). The Google Scholar Experiment: how to index false papers and manipulate bibliometric indicators. *Journal of the Association for Information Science and Technology*, *65*(3), 446-454.

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.

Harzing, A.W. (2007). *Publish or Perish.* Available in http://www.harzing.com/pop.htm

Ley 15/2014, de 16 de septiembre, de racionalización del Sector Público y otras medidas de reforma administrativa. Available in https://www.boe.es/boe/dias/2014/09/17/pdfs/BOE-A-2014-9467.pdf

Mañana Rodríguez, J. (2013). Análisis multidimensional de la especialización en publicaciones de Ciencias Sociales y Humanidades. Doctoral thesis. Available in http://e-archivo.uc3m.es/handle/10016/16964

Michavila, Francisco (Ed.) (2012). *La Universidad española en cifras*. Madrid: CRUE. http://www.crue.org/Publicaciones/Documents/UEC/LA_UNIVERSIDAD_ESPANOLA_EN_CIFRAS.pdf

Sivertsen, G. (2010). A performance indicator based on complete data for the scientific publication output at research institutions. *ISSI Newsletter 6*, 22–28

Torres-Salinas, D., Bordons, M., Giménez-Toledo, E., Delgado-López-Cózar, E., Jiménez-Contreras, E. & Sanz-Casado, E. (2010). Clasificación integrada de revistas científicas (CIRC): propuesta de categorización de las revistas en ciencias sociales y humanas. *El profesional de la información*, *19*(6), 675-684.

Winclawska, B. M. (1996). Polish sociology citation index (principles for creation and the first results). *Scientometrics*, 35(3), 387-391.

NOTES

[1] Law 15/2014, of September 16th, of the rationalization of the public sector and other measurements of administrative reform https://www.boe.es/boe/dias/2014/09/17/pdfs/BOE-A-2014-9467.pdf (Articles 8 and 9)

[2] Today it has become the research group about the Academic Book (ÍLIA): http://ilia.cchs.csic.es

[3] The director of political science affirmed in the assembly of the UNE of November 2014 that the commissions of the CNEAI are sovereign, then they have the decision-making ability in the way they are evaluated.

[4] See the case of Argentina http://www.caicyt-conicet.gov.ar/nucleo-basico-de-revistas-cientificas/

Giménez-Toledo, Elea (2015). La evaluación de la producción científica: breve análisis crítico. *RELIEVE*, *21* (1), art. M2. DOI: 10.7203/relieve.21.1.5160

Author

Giménez-Toledo, Elea (elea.gimenez@cchs.csic.es).
Researcher at CSIC, graduated and PhD in Documentation. Her research focuses on the evaluation processes of scientific activity in the humanities and social sciences. She has co-authored platforms journal evaluation DICE, RESH and CIRC DICE, RESH and CIRC and head of the Spanish part of Latindex. She is coauthor of the ranking of publishers SPI(Scholarly Publisher Indicators). Her postal address is: Centro de Ciencias Humanas y Sociales (CCHS). Consejo Superior de Investigaciones Científicas (CSIC). Albasanz, 26-28. 28037 Madrid (Spain).

To know more / Saber más

ORCID
0000-0001-5425-0003

Google académico

ResearchGate

RELIEVE

Revista **EL**ectrónica de **I**nvestigación y **EV**aluación **E**ducativa
E-Journal of Educational Research, Assessment and Evaluation

[ISSN: 1134-4032]