História, Ciências, Saúde - Manguinhos
ISSN: 0104-5970
hscience@coc.fiocruz.br
Fundação Oswaldo Cruz
Brasil

Mattedi, Marcos Antônio; Spiess, Maiko Rafael
A avaliação da produtividade científica
História, Ciências, Saúde - Manguinhos, vol. 24, núm. 3, julio-septiembre, 2017, pp. 623-643
Fundação Oswaldo Cruz
Rio de Janeiro, Brasil

Available in: http://www.redalyc.org/articulo.oa?id=386153003005

How to cite
Complete issue
More information about this article
Journal's homepage in redalyc.org

Scientific Information System
Network of Scientific Journals from Latin America, the Caribbean, Spain and Portugal
Non-profit academic project, developed under the open access initiative

The evaluation of scientific productivity

Marcos Antônio Mattedi

Professor, Programa de Pós-graduação em Desenvolvimento Regional; coordinator, Núcleo de Estudos da Tecnociência (NET)/Universidade Regional de Blumenau.
Rua Antônio da Veiga,140, Bloco R
89012-900 – Blumenau – SC – Brazil

mattediblu@gmail.com

Maiko Rafael Spiess

Professor, Departamento de Ciências Sociais e Filosofia; coordinator, NET/Universidade Regional de Blumenau.
Rua Antônio da Veiga, 140, Bloco R
89012-900 – Blumenau – SC – Brazil

mspiess@furb.br

Received for publication in December 2015.
Approved for publication in August 2016.

Translated by Diane Grosklaus Whitty.

http://dx.doi.org/10.1590/S0104-59702017000300005

MATTEDI, Marcos Antônio; SPIESS, Maiko Rafael. The evaluation of scientific productivity. *História, Ciências, Saúde – Manguinhos*, Rio de Janeiro, v.24, n.3, jul.-set. 2017. Available at: http://www.scielo.br/hcsm.

Abstract

The paper examines the evaluation of scientific productivity. It analyzes the metrification of the evaluation of scientific production, as well as the historical construction, and current uses of scientific evaluation. It argues that this process contains a paradox: the more that metrics become impersonal, the less they are recognized by scientists. The study is divided into five sections: contextualization of the problematics of scientific evaluation; a description of the main stages in the institutionalization of metrification; an overview of the development of the main evaluation indexes; some examples of the application of these indexes; and analytical consequences and recommendations for the formulation of a new evaluation agenda.

Keywords: evaluation; productivity; metrification; bibliometrics; indicators.

They say the best way to get to know an institution is to discover its obsessions. So if we want to understand the scientific community, we must look at how it evaluates its productivity. The evaluation of productivity is the mechanism by which the scientific community certifies and controls knowledge production. Ubiquitous in science, evaluation generally serves a gamut of purposes; it is used, for example, in awarding grants and other funding, recruiting and promoting scientists, conferring prizes and honors, and so on. Evaluation has become bound up with university rankings and also plays a role in the assessment of research programs, classification of journals, and the judging of the quality of articles and citation patterns. The massification of science communication has turned evaluation into an act of ranking, and ranking into a type of control over scientific activity (Gingras, 2014).

The scientific community is a reputational institution (Whitley, 1982). Its organization was originally based on spontaneous forms of evaluation and thus on implicit hierarchies of the value of knowledge. However, with the massification of scientific activity, these rules were replaced by a formal evaluation system. We know from studies of social stratification that all evaluations endorse classification scales that ascribe prestige to individuals or social groups, as well as to professionals, places, and objects (Constans, Rivoal, 2014). Within the scientific community, prestige and reputation have the effect of structuring the production of scientific knowledge: the more original a scientific discovery, the greater the recognition accorded by the scientific community (Hagstron, 1965). The international literature on evaluation (Donavan, 2007) distinguishes between two strategies: (a) qualitative approaches and (b) quantitative approaches.

The qualitative approach to evaluation is grounded in peer review. Scientific productivity was long limited to peer evaluation, an approach that dates to the latter half of the seventeenth century, when, in 1665, the *Journal des Sçavants*, in France, and *Philosophical Transactions*, published by Britain's Royal Society, instituted the practice of evaluating science communication through specialized review. The practice was gradually extended to the evaluation of the performance of university departments, research programs, laboratories, journals, disciplines, researchers, and so on. However, an excessive increase in the number of evaluations, exacerbated at times by the poor skills of peer reviewers, in tandem with conflicts of interest between them, eventually created problems with peer review (Bornmann, 2008). The practice has consequently come under fire in recent years; growing criticism has been aimed at its central role in scientific evaluation, in light of doubts surrounding both its efficacy in the quality control of papers as well as peer reviewers' subjectivity, corporatism, conservatism, and conflicts of interest (Shatz, 2004; Smith, 2006; Manchikanti et al., 2015).

The quantitative approach came hand in hand with the development of bibliometrics in productivity evaluation. Bibliometrics is a product of the progressive convergence of statistics, sociology, and information technology in the evaluation of researchers, groups, and institutions. More precisely, it entails procedures that contribute to productivity evaluation based on number of publications, the prestige of the publishing journals, and citation patterns (Académie des Sciences, 2011). This approach has gained increasing ground because it affords distilled, factual information on the dynamics of the scientific community. While bibliometrics may produce serious distortions when used in isolation,

its application in the evaluation of scientific productivity has fueled much excitement and transformed scientific articles into a central factor in this evaluation.

The gradual replacement of a reliance on peer review with a reliance on metrics in such matters as recruitment, funding policy, and institutional evaluation (Gingras, 2014) has been accompanied by a profound change in the scientific community's understanding of itself and organization. On the one hand, this process has reflected the ebbing of the epistemological description of science as a rational cognitive activity (Popper, 2006) and the establishment of a sociological view of science as a social activity (Kuhn, 1989); on the other, it has reflected a change of a sociological nature, as the massification of science has transformed knowledge production into a collective enterprise with rising investments (Price, 1963). Thus, the development of monitoring and scientific direction has been associated with two processes: evaluation and funding (Whitley, 2000). The practical relation between these two processes contains a paradox: the greater the scientific excellence of knowledge (originality), the lower the social accessibility of this knowledge (understanding).

It can be argued that this emphasis on the quantification of scientific activity has countless unintended consequences. While this "numericizing" (Desrosières, 1998) provides ways of simplifying and objectifying social facts (the productivity of scientists and institutions, their collaborative relationships, or the dynamics within a field of knowledge), it becomes an end in itself. In many cases, rather than guiding science policy and resource distribution, the logic of productivity evaluation leads to the phenomenon that has been described as "accelerated academy" or "productivity culture" and that we will call "productivism" in these pages, in reference to the Brazilian literature on the concept (*produtivismo*). Productivism opens the way for such practices as spurious self-citation, so-called salami publication (slicing results from one project into a number of articles), and growing incidents of plagiarism and scientific retraction (Castiel, Sanz-Valero, Red Mei-Cyted, 2007; De Bellis, 2014; Sguissardi, Silva Júnior, 2009). In other words, the centrality of these forms of evaluation seems to put pressure on scientists in a way that introduces irregularities and anomalies into their traditional practices.

Recognition of the limits of scientific evaluation and the impact of productivism has given rise to criticisms and reactions, often in the form of manifestos released by researchers, institutions, and associations. Examples include the San Francisco Declaration on Research Assessment (Dora), published in 2012, which calls into question the correlation between metrics like journal impact factor and researcher merit (Alberts, 2013); the Leiden Manifesto (2015), which is based on the premise that evaluations are "usually well intentioned, not always well informed, often ill applied" (Hicks et al., 2015, p.429) and which suggests ten new principles applicable to research evaluations; and the Force11 Manifesto: Building the Future for Research Communications and e-Scholarship (2011), which reflects on new ways of publishing science. Their common thread is that they all point to a breakdown in the prevailing model for developing and applying scientific evaluation.

Lastly, the current state of scientific evaluation can be criticized not only from a practical perspective but also from an ethical one (Furner, 2014). All forms of evaluation are arbitrary; they represent different views of scientific activity and choices regarding resource distribution, merit, and visibility among researchers and institutions. Yet the systematization and use of certain forms of evaluation favors certain practices while concomitantly discouraging

others. This homogenizes the practice and stifles dissidence; in other words, it compels certain behaviors, values, and priorities. Moreover, it creates a distributive justice problem: when certain professional practices and profiles are more valued, they tend to foster the accumulation of resources and prestige. Evaluation thus presents additional obstacles and challenges for fledgling researchers, those working with neglected topics, or those located on the periphery of the system.

It is our view that metrification is a combined product of these sociotechnical processes. More precisely, metrification represents the cognitive, normative operation that is employed in an effort to transform productivity evaluation into an unbiased, reliable enterprise (Porter, 1995). It is about designing and applying objective measurement instruments, that is, about standardizing evaluation measures. The process has to do both with the formation of data production experts and with the establishment of a social foundation that endows these measures with authority. We argue here, based on an analysis of the formation and development of scientific evaluation, that metrification represents the transformation of the paper into the main product of scientific activity. This transformation has three implications, which we will call: (a) paper-centrism; (b) productivism; (c) mimesis.

With this characterization of the process of the metrification of evaluation in mind, our argument is laid out in the next sections, which explore: (1) the historical institutionalization of metrification, highlighting the main stages of this process; (2) scientific evaluation indicators and their growing diversity and complexity, with an examination of some of the main productivity and relatedness indexes; and (3) current practice in scientific evaluation, with a focus on its possible uses and a look at its pros and cons. In the last section, we offer proposals for overcoming the dilemmas raised by the current debate over the evaluation of scientific activity. In short, based on an analysis of the process of the institutionalization of metrics and their application, we portray the emergence and consolidation of the metrification of scientific evaluation.

The institutionalization of scientific evaluation

Interest in generating, communicating, and applying knowledge has kept step with the formation and development of the scientific community (Mattedi, Spiess, 2010). On the one hand, it has been related to an increase in the number of publications and the perception that this process can be described scientifically; on the other, it has been accompanied by a steady rise in investments and the re-dimensioning of the academic and institutional limits on research. Accordingly, the need for evaluation ties in with the application of statistical methods to scientific literature but also with the need for management and control. Put more precisely, it ties in with the need to understand the meaning, features, and differences of the combined outcome of the individual production of scientists in different disciplines. The institutionalization of scientific evaluation can be divided into three main stages: (a) design; (b) stabilization; (c) dissemination.

The institutionalization of scientific evaluation began in the early decades of the twentieth century, when quantitative methods of analysis were first applied to publications, authors, and bibliographic references. It is grounded in the scientific article, which is the standard form of

science communication. Using scientific articles offers some advantages: they have an easily identifiable author (or co-authors) and a passive bibliography and they are unchangeable once they have been published on paper. In other words, the initial development of evaluation saw the progressive extension of statistical data handling methods, with the scientific article as its empirical basis. This brought the emergence of a reductionist attitude, whereby the evaluation of science was reduced to an easily collectible, observable dimension, with countable units, in perfect Positivist style (De Bellis, 2014). The idea was to understand science as a social institution by analyzing its members and their output.

This approach took root in the form of bibliometric laws. One of the earliest contributions, published in 1926, came from chemist Alfred J. Lotka. Lotka's law (or Inverse Square Law of Scientific Productivity) states that the number of authors who make *n* contributions in a given field of knowledge is approximately $1/n^2$ of those who make only one contribution; in other words, in a given field of knowledge, the proportion of authors who contribute with only one publication is around 60% of all authors in the field (Coile, 1977). Along similar lines, Bradford's law, originally formulated in 1934, is aimed at journals and tries to "determine the core and areas of dispersion in a certain subject within one same set of journals" (Vanti, 2002, p.153). Lastly, Zipf's law analyzes word frequency in texts to arrive at an ordering of the terms most often used in a scientific field.

These laws were virtually ignored by the scientific community until Derek de Solla Price re-introduced the discussion proposed by Lotka in his books on the growth of science (McRoberts, McRoberts, 1982). Their influence and interpretation have varied greatly since then. In conceptual and methodological terms, their reproducibility has been tested in various fields of knowledge and databases, with ambiguous results (Pinheiro, 1983; Urbizagástegui Alvarado, 2002). More specifically, their validity and scope in different fields of knowledge (for example, the humanities) and different contexts have been called into question, and doubts exist about the universality of the laws and their implications for the organization of scientific activity. After all, can scientific activity be understood solely by measuring its output?

In summary, Lotka, Bradford, and Zipf based their work on the statistical analysis of the activities of the individuals comprising the scientific community. They centered on the notion of authorship, as expressed through publication in specialized journals. On the one hand, this is an overly narrow view of the functions of science, as suggested by John D. Bernal (1967), in that it takes into account nothing but the final product of scientific activity while excluding gray literature, that is, the transfer of tacit knowledge, which is the everyday routine of science. On the other hand, this approach is a manifestation of the Matthew Effect, as formulated by Robert K. Merton (1968), where by eminent researchers tend to get more credit than their less well-known colleagues and therefore enjoy more prestige, access to funding, and visibility. These limitations not withstanding, Lotka's and Bradford's pioneer contributions continue to wield much influence in scientific evaluation strategies.

After World War II, the proliferation of scientific literature and the challenge of assessing its relevance became an obstacle to new research. Inspired by the legal world's Shepard's Citations, Eugene Garfield proposed the Science Citation Index (SCI®). As editor of the journal *American Documentation*, Garfield asked William Adair (1955), a vice-president at Shepards Company, to write a paper describing how the legal tool worked. Some months

later, writing in the journal *Science*, Garfield (1955) suggested that a database be created that would allow cited articles to be used to help locate other articles. His proposal derived from an intuitive sense that there was a conceptual link between citing article and cited article (Gingras, 2014). The SCI offered a new image of scientific literature, just as a telephone book fashions an image of a city's residents (Wouters, 1999).

In 1959, confident in the feasibility of his proposal, Garfield founded the Institute for Scientific Information (ISI), replacing the former Eugene Garfield Associates, Inc.; while its name gave the impression that the new company was a governmental agency, its structure allowed it to compete with not-for-profit organizations. In 1961, Garfield received a $300,000 grant from the National Science Foundation (NSF) and the National Institutes of Health (NIH) to devise an automatic index for disseminating and retrieving information (Garfield, 2007). Following its initial support, the NIH was barred from funding businesses. The NSF then negotiated a contract with ISI for 1,000 copies of a Genetics Citation Index (GCI). The SCI provided a new representation of science in terms of the production and consumption of scientific information.

The initial idea behind the SCI was to make it easier to locate literature. As stated by Garfield (2007, p.65): "The SCI's multidisciplinary database has two purposes: first, to identify what each scientist has published, and second, where and how often the articles by that scientist are cited." The goal of the SCI was thus to make access fast and automatic by reducing complex scientific language to a set of manageable units, that is, to a specific set of metatextual relationships derived by linking journal articles to bibliographic references. Interest in analyzing the SCI comes from the fact that, as a tool of productivism, the SCI effectively obscures the content of the literature by focusing on its formal properties. The process established a new representation of science, different from the description of scientists' cognitive or behavioral processes (Wouters, 1999).

The SCI quickly caught attention in the sociology of science, because it enabled investigation of the workings of the scientific community. By aggregating the properties of publications, references, and citations, it made it possible to test hypotheses derived from the Mertonian understanding of scientists' behavior (Mattedi, 2006), triggering first clashes and later rapprochement between the Columbia School, which followed Merton, and the Philadelphia School, influenced by Price, with quantitative accents (Elkana et al., 1978; Wouters, 1999). Solla Price was the first author to use the index created by the SCI; in a quantitative examination of data on the development of science, Price (1978a) found a skewed pattern in the distribution of scientific production, meaning that the growth of scientific information was much faster than other social phenomena, yet quite similar to other phenomena observable in natural contexts. This relation earned the name Price's Law, according to which 25% of scientific authors account for 75% of published articles (Price, 1963).

As electronics and computer science moved forward, new possibilities were opened for the measurement and analysis of scientific production. After World War II, steady technological advances, like integrated circuits and microprocessors, gradually made computers smaller while augmenting their processing capacity, putting microcomputers on the market in the 1970s and personal computers in the 1980s (Mowery, Rosenberg, 2005). The dissemination

of these technologies occasioned a veritable revolution that reached the administrative and productive realms of Western societies and brought economies of time and scale, so that large amounts of data could be analyzed and processed in little time. These applied technologies also made major changes in the area of scientific evaluation possible.

It should be pointed out that these computational technologies allowed researchers and those working on initiatives like the SCI to implement increasingly complex forms of statistical analysis. This correlation between available technology and the evaluation of scientific activity was highlighted by Solla Price (1978b) in the editorial he wrote for the first issue of *Scientometrics*. One example of how it took firm hold is impact measurement. Originally proposed by Garfield in 1955, the impact factor is based not only on the number of individual publications by an author but also on their influence or the importance of a journal within a given field. The factor was determined by analyzing all periodicals covered by the SCI; in 1969, this meant analyzing thousands of references in over 2,200 journals (Bensman, 2007), a task that would have been impossible without the aid of computer technologies.

More recently, with the popularization of personal computers and the Internet, science communication has become ever more electronically based. This at first occasioned the multiplication and fragmentation of databases, according to field and geographic and linguistic scope. New forms of access to scientific knowledge gradually surfaced, engendering new practices and problems in the evaluation of scientific activity. One of the most important shifts occurred in 1992, when the Thomson Corporation purchased the ISI and its products (SCI and similar indexes) and the Web of Science® was created (De Bellis, 2014). This was the moment when conceptual advances and technical means began to complement each other and became cemented in the form of a business model, giving birth to other bases, like Scopus, Elsevier, and Google Scholar.

The ties between Information and Communication Technologies (ICT) and scientific activity brought profound change to how scientific activity is communicated and evaluated. More precisely, a new benchmark in evaluation was set by the publication of articles online or in hybrid journals, listed in myriad other articles and potentially relevant digital objects, linked in real time and through hyperlinks. The combined effect of this process was the materialization of a kaleidoscope of evaluation methods, such as infometrics, scientometrics, cybermetrics, webometrics, influmetrics, digimetrics, and other neologisms referring to this process. At the same time, a panoply of journals, professional organizations, conferences, prizes, syllabuses, and research centers related to scientific evaluation, involving investment funds, corporations, and universities, came into being (Cronin, 2014).

Accordingly, the institutionalization of the evaluation culture expresses the institutionalization of the article and the incorporation of references as parameters for evaluating scientific activity. The adoption of productivity measures (publication counts) and impact indicators (citations counts) is indicative of a shift from little science to big science (Price, 1963). It is also an expression of cooperative ties and rivalry between Merton, Bernal, Price, and Garfield (Elkana et al., 1978). Lastly, it reflects a shift in how information is communicated, from paper-based to e-based. The Internet, in tandem with the expansion of statistical methods, made it possible to handle enormous databases, and the consequences were more sophisticated measurement methods and the monitoring of various facets of science

communication in general and scientific evaluation in particular, at the macro (country), meso (discipline), and micro (program) levels.

The proliferation of scientific evaluation indexes

The past two decades have seen a drive to create new indicators. On the one hand, this has been bound up with growing demand for information on the part of researchers, funding and evaluation agencies, editors, and journals, while on the other, it stems from the availability of large international databases and enhanced statistical, sociological, and computational abilities. The result has been the proliferation of indexes like those based on indexed publications, field of specialization, visibility and diffusion, institutional and international collaboration, and indicators of use and recommendation. These metrics vary according to type of count, nature of calculation, and standard of measurement. According to Callon, Courtial, and Penan (1995), they can be divided into two main groups: (a) productivity indexes; (b) relatedness indexes.

Productivity indexes are based on the premise that science and technology are productive activities that can be measured and understood in terms of input and output, where funding, material, and labor enter in and the results of scientific activity come out, in the form of articles, patents, instruments, and trained professionals. This means that the task of evaluating scientific activity is to measure the volume of production and its impact in a given field of knowledge, so as to ascertain its dynamism and evolution, along with individual researcher contributions and productivity. This is an intrinsically numerical, statistical form of evaluation that describes and analyzes scientific activity in instrumental terms, that is, in terms of performance and impact. These measurements can be divided into two types: publication count and citations count.

Publication count is the simplest index of scientific production (Callon, Courtial, Penan, 1995). It is grounded in the principle that the activity of researchers or groups within a field of knowledge, specialty, or geographical region can be measured by identifying and counting the number of articles published in academic journals. From the perspective of individual researchers, it enables an analysis of the quantitative evolution of their production and a comparison with the résumés of other professionals. Publication count is therefore the starting point and mechanism for verifying certain proposals about author ability and contribution, such as Lotka's Law. An aggregate analysis of these indicators also helps measure productivity rates within a discipline or the participation of an institution or country in overall scientific production in a given timeframe.

On the one hand, this type of evaluation intends to rationalize resource allocation and public-policy making in science, within a context of growing competition (Velho, 1985; Leta, 2011). Its use has been systematized and disseminated through initiatives like the *Frascati Manual*, first published in 1963, which sought to establish standardized forms of scientific evaluation. On the other hand, however, it is grounded in a reductionist view of scientific activity: when the focus is on specific output (written knowledge in the form of a scientific article), various other results and output are rendered invisible. Moreover, it naturalizes the image of science as a cumulative activity subject to the laws of statistics.

In this sense, publication count by author, institution, or country may seem self-evident or reified and thus appear to forego the need for problematization or contextualization.

Another common indicator in scientific evaluation is citation count. The basic premise here is simple: the most influential articles and scientists are the most cited, and the number of citations indicates reception (Glänzel, 2008). Existing links between documents thus make it possible to assess how useful an article is to another researcher and thus estimate its importance. After all, there is a correlation between high citation indexes and peer judgments about the scientific excellence of contributions (Garfield, 1979). This means that the greater the number of citations, the more important the article and the scientist. So when an article is cited, two hypotheses can be raised about its importance: (a) it is visible enough to serve as a reference; (b) it has an impact that can only be measured by the document itself (Callon, Courtial, Penan, 1995). Citation count enables an evaluation of degree of utilization and can thus measure the impact of articles and journals.

The operationalization of the method for counting citations is straightforward and based on a paper's formal elements: author, institution, title, journal, location, number of pages, date, bibliographic references, and so on. Since almost all scientific documents include references, then the total number of articles, communications, letters, reports, etc. published in science journals can be cross-referenced with footnotes and bibliographic references. The resultant database can be used to count all references to journal J in year Y, yielding an impact factor (Garfield, 1972). The continual sophistication of these has led to the design of software like Publish or Perish (PoP), released in 2006 by Microsoft Academic Search. In addition to furnishing simple statistics (document count, citation count, etc.), PoP calculates individual citation metrics like the h-index, formulated by Jorge E. Hirsch, and the g-index, devised by Leo Egghe (Harzing, 2011).

Much controversy has surrounded the adoption of productivity indicators in scientific evaluation (Cozzens, 1981; Leydesdorff, 2001), and no consensus has been reached about how best to apply them. After all, while one group of researchers wonders what is actually measured by citation count, another group is concerned about what it might not measure. So although productivity indexes are seen as powerful instruments for charting the intellectual impact of scientists, journals, disciplines, and programs, the validity of these data is called into question based on the suitability of databases in research evaluation. For example, output in a narrow field that is not published in English or in book form is often times not captured, as we see in the social sciences. To put it another way, productivity indicators fail to identify and even underestimate those in the scientific community who are not already recognized.

Relatedness indexes were devised to detect links between bibliographic elements. They describe the degree of similarity or difference between documents, authors, journals, and concepts and provide a measure of the strength of the links. This means it is possible to position and group interactions between these elements. Bibliographic references are no longer considered isolated, disconnected entities but a whole, based on intrinsic relationship rules like co-word analysis and on software like T-LAB. Relatedness indexes differ in their content analysis: first-generation do not enter into content analysis (co-authorship and citation networks), but second-generation do (word co-occurrence, co-classification of publications, and co-citations) (Callon, Courtial, Penan, 1995).

The main first-generation relatedness indicator is co-authorship analysis, which is used to determine cooperation between institutions or research teams. The basic idea is that the number of joint articles produced expresses cooperation in research activities, meaning that co-authorship reflects the professionalization and specialization of a community of authors in terms of collaboration and funding. Thus, the greater the number of co-signed articles, the greater the funding these authors receive for their research. The main elements measured are number of authors, their ranking, and principal investigators, as well as heterogeneity among academic and industry researchers. Co-authorship analysis thus maps networks of cooperation among researchers and sheds light on the dynamics of the scientific community.

Beyond the focus on authorship and citations, other relatedness indexes aim to capture the dynamics and density of contributions within a given topic or specialty. One example is the analysis of word co-occurrence, a statistical method for analyzing pairs of words or phrases in order to identify recurring patterns that reflect a link between concepts within a corpus of texts. Also known as Leximappe, the method was developed in the 1980s by the Centre de Sociologie de l'Innovation, of the École Nationale Supérieure des Mines de Paris, and the Centre National de La Recherche Scientifique (CNRS) (He, 1999). It provides a representation of a network and the links between concepts, issues, and ideas. In short, its purpose is to create relatedness indicators that make it possible to chart and understand the evolution of science and technology. Even though it is a statistical method, its goal is thus intrinsically policy oriented (Courtial, Law, 1989).

Similarly, relationships between science journals can be analyzed. The starting point is citation count by journal, an approach first suggested by Katherine W. McCain (1990). This relatedness index shows which publications are related to the same topics within a given field or specialty. It also makes it possible to build a map or network wherein the number of cross-citations denotes the strength of links between journals and, indirectly, the density of a discipline. This analytical approach seeks to complement analyses of a basically quantitative nature, such as classifications according to total articles published or impact factor, the latter based on the ratio of recent citations of articles published in a journal and the total number of articles published in a given timeframe.

Another second-generation relatedness indicator is author co-citation analysis (ACA), which involves pairing data or co-cited authors through statistical methods such as cluster analysis, multidimensional scaling, and factor analysis. Designed by Howard White and Bewar Griffith in 1981, the method is rooted in the assumption that when two citations are found in the same text, their relationship indicates proximity of content, suggesting that the number of references common to two or more texts is an indicator of cognitive proximity. ACA makes it possible to form classes of fields and therefore to devise clusters (Andrés, 2009). ACA has been taken to different levels of aggregation, such as journal, author, and topic, in order to examine and sketch out the structure of the community of researchers and disciplines.

In summary, differences and similarities between productivity and relatedness indicators show that scientific evaluation has many uses. The move from productivity indicators to relatedness indicators reflects a progressive shift of interest away from inter-bibliographic elements to intra-bibliographic ones, signaling the development of methodologies that respond to the automatic indexing trend. At the same time, it also reflects the entrance of

new actors and a shift in the discussion away from the United States toward Europe, a trend indicative of new needs. The combined effect of these two changes has not only been the proliferation of indicators but also the replacement of deliberative methods with quantitative ones (De Bellis, 2014; Cronin, Sugimoto, 2014). In practical terms, this means that it has become more challenging to choose measurement tools appropriate to scientific activity.

The applicability of indicators in the evaluation of scientific activity

Quantitative indicators are not just scientific products (Van Raan, 2004); they also serve as tools in evaluating, regulating, and drawing up policy (Narin, 1976). They were first applied to the evaluation of productivity in the late nineteenth century (Godin, 2009) and have often been linked to research funding. Bibliometric indicators are currently one of the key tools in evaluating the scientific capacity of individuals or institutions (Hicks et al., 2015). One of the main indexes is the impact factor, formulated by Eugene Garfield in 1952 to ascertain the quality of science journals, but there are other, lesser-known indicators, such as the immediacy index, prestige factor, and usable factor. This section will examine the application of indicators to three levels in the analysis of scientific activity: micro (researchers); meso (journals); and macro (institutions).

Researcher evaluation is probably the topic that has drawn greatest attention from the scientific community over the last decade. This process has accompanied the massification of scientific activity, but it also relates to the limitations of peer review. Researchers have consequently grown more concerned with boosting their number of publications and citations. There are two ways of performing this calculation: (a) manually, that is, through the individual analysis of articles; (b) automatically, using database information, such as Publish or Perish, to assess members of a meeting, the editorial board of a journal, attendees at a scientific event, and so on. Two good examples are the h-index and the g-index.

The h-index was formulated by Jorge E. Hirsch in 2005 to measure an individual scientist's productivity (Hirsch, 2005). The index combines quantity measurements (publications) with impact measurements (citations). In Hirsch's words: "A scientist has index h if h of his or her N_p articles have at least h citations each and the other ($N_p - h$) articles have $\leq h$ citations each" (p.16569). Therefore, a scientist has an h-index of fifty if he or she wrote fifty articles that have at least fifty citations each. If the h-index is an indicator of recognition on the part of the scientific community, forging a successful academic career is about obtaining increasingly greater recognitionin order to break through h thresholds (Grupo Scimago, 2006). However, this indicator distorts individual evaluation because it does not allow for comparisons between disciplines (reference and article counts) and especially because it penalizes scientists who publish selectively (Costas, Bordons, 2007).

The g-index, designed by Leo Egghe in 2006, was a proposal to enhance the logic applied to the h-index. Given a set of articles ranked in descending order of number of citations received, the g-indexis the largest g value in which the first g articles combined received at least g^2 citations (Egghe, 2006). For example, a researcher who has four publications with citation counts of five, three, one, and one has a g-index of three. The g-index thus renders the impact difference between authors more visible, while it also evinces the importance of

the author's main articles. A larger g-index therefore represents more and better articles (Tol, 2008). However, since the g-index, like the h-index, is represented by a whole number, many authors can score the same, making it harder to differentiate between them (Huang, Chi, 2008).

Taken together, the h-index and g-index are quantitative indicators that aim to differentiate scientific authors in terms of the impact of their contributions. Both are relatively new ways of evaluating researcher output, along with other measurements like the a-index, 4-index, and r-index, which are gaining ground as tools for ranking professionals within a field, institution, or department (Schreiber, 2008; Selek, Saleh, 2014). They seek to offer alternatives to evaluation methods based on raw production counts, which can lead to the distortions of productivism, yielding freeriding researchers and extremely low-quality papers. Yet they have so far failed to overcome the distortions derived from a researcher's position, from his or her access to resources and collaboration networks, and even from how researchers use their knowledge of productivity metrics in their favor.

The application of metrics to science journals establishes an equivalence between evaluation and classification. If there has always been a tacit hierarchy of journals within each discipline, this process becomes institutionalized with the introduction of quantitative measures (Gingras, 2009). For example, the impact factor has an effect on a journal's reputation in that a higher factor will draw more submissions and better articles. The evaluation thus becomes a way of assigning prestige, in turn impacting the structure of science communication. A number of rankings can serve to illustrate this, for example, France's Listes de Revues Sciences Humaines et Sociales da Agence d'Évaluation de la Recherche et de l'Enseignement Supérieur (Aeres) or Excellence in Research for Australia, published by the Australian Research Council (ARC). However, we will look at only three: (a) *Journal Citation Reports* (JCR); (b) the European Reference Index; (c) and Qualis Capes.

The JCR is a bibliometric product published by the Thomson Reuters (2015) group, which offers ways "to critically evaluate the world's leading journals"(s.p.). The JCR derives from the Science Citation Index originally proposed by the ISI, but it obeys a different logic: if the focus was originally on the author, the key to organizing and ranking data with the JCR is a list of journals and their output (Garfield, 2007). Analysis of data compiled in the JCR yields an impact factor, immediacy index (number of articles from a journal cited in the same year as their publication), and other, similar indicators. The tool thus intends to provide criteria for judging the relevance of a publication, especially as an aid to bibliographic research (Pendlebury, Adams, 2012). In practice, however, it has gradually fostered differentiation and has concentrated importance among publications.

The European Reference Index for the Humanities (Erih Plus) is an index of journals in the humanities. It was developed by the European Science Foundation (ESF) in 2005 and transferred to the Norwegian Social Science Data Service (NSD) in 2014. All journals encompassed by Erih Plus are ranked into three categories, according to scope and public: A (international publications with a strong reputation among researchers); B (international publications with a good reputation among researchers); C (regional reputation, of local importance). But between its conception and execution, the ranking became a way to attribute quality that ascribes greater value to those publishing in category A (Editorial, 2009).

Qualis Capes employs a set of procedures to stratify Brazilian scientific production. It was designed to assess and furnish a list that ranks graduate-level output. Publication quality is gauged indirectly, based on an analysis of journal quality. As stated by Capes (2014, s.p.): "The classification of journals is undertaken by evaluation areas and is updated annually. Journals are ranked into strata that indicate quality: A1, A2, B1, B2, B3, B4, B5, C, wherein A1 is the highest and C is equal to zero." Based on their impact factor, A1 and A2 journals display "international excellence"; B1 and B2, "national excellence"; B3, B4, and B5, "average relevance"; and C, "low relevance" (Ferreira, Antoneli, Briones, 2013), thus running counter to international recommendations like those issued by the San Francisco Declaration on Research Assessment. In practical terms, this evaluation and classification system affects an individual researcher's publication decisions as well as the editorial procedures and quality processes employed by journals (Frigeri, Monteiro, 2014); it can even affect chances of obtaining funding (Silva, 2009).

The use of bibliometric tools to evaluate laboratory or research programs has been a matter of some controversy. While the scientific productivity of a scientific organization can be viewed as a product of investments, there is no protocol about how to go about it (Okrasa, 1987). Nevertheless, a number of initiatives have been undertaken to assess and, above all, justify funding allocations, such as the *Frascati Manual* (OCDE, 2007). This has sparked clashes since there is no linear model, in administrative, economic, or bibliometric terms (Godin, 2009), that enables the comparison of organizations with very distinct institutional, administrative, and financial profiles. Although there is no algorithm for evaluating the performance of a body of researchers and serving as a benchmark indicative of scientific productivity, one can take the case of research programs and universities.

Since the seminal study by Martin and Irvine (1983), the evaluation of research groups (programs, laboratories, schools etc.) has been based on a publication's international influence. This means that scientists who have something important to say strive vigorously to publish their discoveries in international journals (Vinkle, 2010). The extension of this assumption to the study of research groups can be illustrated by examining the publication practices of the Leiden School (De Bellis, 2014), which grew out of the Centre for Science and Technology Studies at Leiden University. We can take the example of the CPP/FCSm Indicator, which associates citations per publication with average score in the field of the citation (Van Raan, 2004). This indicator enables comparison of the citations of all articles published across all journals with the worldwide average, thereby establishing an institute's productivity.

University rankings offer a way to evaluate the quality and relevance of university teaching and research. Two rival metrics currently dominate the world stage: the Academic Ranking of World Universities (ARWU), produced by Shanghai Jiao Tong University in China since 2003, and the Times Higher Education World University Rankings, created in 2004. Both are based on indicators that evaluate performance in terms of teaching and faculty quality, citations, recognized indexes, per capita performance, internationalization, and industry investment, with each area assigned its own weight in calculations. The Ranking Universitário Folha, created in 2012, uses a similar rationale to analyze Brazilian universities, based on five indicators: research, internationalization, innovation, teaching, and market.

In a context of massification of higher education and scientific research and of rising competition among institutions, such rankings of research groups, institutes, and universities serve to guide the allocation of funding and personal prestige and to regulate competition between institutions for the brightest, most promising students and researchers (Altbach, 2006). The problem with this type of evaluation does not lie in its principles but in practice. Various rankings currently tend to present consistent results for the best-ranked institutions but not for the ones that score lower (Saisana, D'Hombres, 2008; Usher, Savino, 2007). This type of evaluation thus seems to certify institutions that already enjoy prestige, display a classic scientific orientation, and have greater access to funding, while at the same time attributing negative evaluations to institutional profiles lying on the periphery, either geographically or in terms of field.

An examination of the application of indicators to the evaluation of scientific productivity shows that there is more information on scientific production than on their application. It can therefore be said that a productive researcher today is not just a scientist who publishes but a researcher who publishes in certain journals and a specific number of articles per year; a journal that enjoys credibility is not just one that is recognized by members of a given scientific discipline but one that performs well in rankings. Increased productivity on the part of a scientific organization is always accompanied by the distancing of the regional community. What should be constructed from data is not a number but a pattern of meaningful elements that render the process of change visible. It can thus be said that the asymmetry between the quantity of metrics and their application indicates that these practices are being adapted to the criteria (Gingras, 2014).

Contributions to a new agenda in scientific evaluation

The metrification of scientific evaluation is a combined outcome of the progressive integration of statistics, sociology, and information technology. It reflects the refinement of statistical tools, the scientific community's pattern of organization, and technological supports. It also derives from pioneer contributions by Lotka, Bradford, and Zipf as well as Merton, Bernal, Price, and Garfield, and from the need to manage and control scientific activity. This process of quantifying information, massifying science communication, and developing databases has guided scientific production and reshaped its meaning. Metrification thus emerged in a very specific context but ultimately spread throughout the scientific community. What this examination of the process has revealed is that there is more information about the production of indicators than about their application.

Metrification presents an intriguing paradox. The more the evaluation of scientific productivity is refined in technical terms, the lower scientists' confidence in these tools. Put more precisely, the greater the objectivity of a tool, the lower its credibility. The key to interpreting this paradox is recognizing that in metrification, objectivity does not stem from the knowledge gathered over the course of one's career but from the application of rules unknown to the scientific community. The proliferation of metrics thus suggests that each group ends up developing its own parameters and indicators to justify its own scientific practices. That is why in most cases, metrics are accepted when convenient and snubbed

when unfavorable. When metrification appeals to the impartiality of numbers, it standardizes local competence into general rules, transforming one pattern of science communication into a parameter for evaluating all scientific production.

The controversies fueled by this paradox find expression in various types of institutional reactions. From a methodological perspective, we see an ongoing process of refinement and multiplication of evaluation techniques, indicative of the uncertain nature of these tools. From a political perspective, we also see certain fields of knowledge resisting the tools, as is the case of the humanities in general and the social sciences in particular. In this sense, manifestos like the Force11 Manifesto: Building the Future for Research Communications and e-Scholarship (2011), the San Francisco Declaration on Research Assessment (2012), and the Leiden Manifesto (2015), along with their ensuing debates, are only the most visible manifestations of a generalized sense of ill-ease within the scientific community. The paradox thus seems to breed conflicts of interest between the public and private sides of science.

Furthermore, the strategy of quantifying individual productivity and evaluating aggregate data (or "large numbers") has generated countless distortions in scientific activity, from a normative perspective. In other words, emphasis has shifted to the output of a researcher, group, or institution in detriment to the values traditionally associated with the scientific community, like intellectual autonomy and political independence. As a result, the unintended consequences of the metrics of science include frequent cases of plagiarism (and ergo retraction), self-citation, redundant publications, undue attribution of authorship, and freeriding researchers. In short, irregularities have arisen in peer communication and in the definition of science and technology policy because evaluation has centered on articles and productivity measures and because of the issues related to the assignment of funding and prestige.

The metrification of productivity evaluation is thus the history of how the scientific article has become an expression of scientific activity. A new evaluation agenda must overcome three obstacles engendered by metrification:

(a) Paper-centrism: a scientific article should not be considered the focus of scientific evaluation.

(b) Productivism: a good researcher is not just someone who scores well on existing rankings.

(c) Mimesis: international recognition cannot be considered a benchmark for certifying knowledge.

A new evaluation agenda therefore requires an "anti-reductionist" posture: scientific evaluation cannot be reduced to an analysis of scientific literature; the usefulness of an article cannot be reduced to its visibility within the scientific community; and the excellence of scientific production cannot be reduced to international similarity.

This agenda can be developed by investigating the causes and impacts of new phenomena related to the everyday routine of scientific activity. This investigation can, for example, explore transformations in the notion of authorship and credit among scientists (through the division of authorship among large research groups) as well as crowd or networked science, where professionals and amateurs work together on the same problem, often coordinated

via complex computer platforms. Additionally, it is possible to research such cases as the organizations of patients and families who become "lay specialists," collaborating with physicians and scientists in the discovery of new treatments and redefining the boundaries of scientific production and the role of the article as a method of dissemination.

Two strategies should be used to re-assess the evaluation of scientific activity. First, it is necessary to understand the historical formation of the means of scientific evaluation and the process of the metrification of science, its epistemological and political implications, and, especially, its current limitations. Second, new modalities in the production of scientific knowledge must also be understood in order to propose new means of evaluation. If existing metrics and indicators have already proven limited in their ability to accompany "normal science," it is obvious that their evaluative ability will be even narrower in cases where the attribution of authorship and forms of dissemination are radically new or heterodox. It becomes necessary to redefine the problem and build new forms and meanings in the metrics of scientific evaluation, moving beyond today's productivist obsession in science.

Doing away with this cognitive monopoly should progressively modify the evaluation of scientific productivity. The control of scientific quality has long been restricted to evaluation ex ante: knowledge is first certified within the scientific community through peer evaluation and then it spills over into society. However, with the rise of the Internet and as scientists have lost their monopoly over knowledge production, new experiences are being tested out, such as post-publication evaluation. The latter, which keeps step with initiatives like Wikipedia and other types of hypertexts, assesses knowledge expost, through post-publication peer review via sites like PubPeer. This means that knowledge becomes public and open to any type of contribution, conjoining the scientific community's internal evaluation with society's external evaluation.

REFERENCES

ACADÉMIE DES SCIENCES.
Du bon usage de la bibliometrie pour l'evaluation individuelle des chercheurs: rapport remis le 17 janvier 2011 à Madame la Ministre de l'Enseignement Supérieur et de la Recherche. Disponível em: http://www.academie-sciences.fr/archivage_site/activite/rapport/avis170111_synthese.pdf. Acesso em: 1 dez. 2016. 2011.

ADAIR, William.
Citation indexes for scientific literature? *American Documentation*, v.6, n.1, p.31-32. 1955.

ALBERTS, Bruce.
Impact factor distortions, *Science*, v.340, n.6134, p.787. 2013.

ALTBACH, Philip G.
International higher education: reflections on policy and practice. Chestnut Hill: Boston College Center for International Higher Education. 2006.

ANDRÉS, Ana.
Measuring academic research: how to undertake a bibliometric study. Oxford, Cambridge, Nova Déli: Chandos Publishing. 2009.

BERNAL, John D.
The social function of science. Cambridge: The MIT Press. 1967.

BORNMANN, Lutz.
Scientific peer review: an analysis of the peer review process from the perspective of sociology of science theories. *Human Architecture: Journal of the Sociology of Self-knowledge*, v.6, n.2. p.23-38. 2008.

BENSMAN, Stephen J.
Garfield and the impact factor. *Annual Review of Information Science and Technology*, v.41, n.1, p.93-155. 2007.

CALLON, Michel; COURTIAL, Jean-Pierre; PENAN, Hervé.

Cienciometria: el estudio cuantitativo de la actividad científica: de La bibliometría a la vigilancia tecnológica. Oviedo: Trea. 1995.

CAPES.
Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. *Classificação da produção intelectual*. Disponível em: http://www.capes.gov.br/avaliacao/instrumentos-de-apoio/classificacao-da-producao-intelectual. Acesso em: 1 dez. 2016. 2014.

CASTIEL, Luis David; SANZ-VALERO, Javier; RED MEI-CYTED.
Entre fetichismo e sobrevivência: o artigo científico é uma mercadoria acadêmica? *Cadernos de Saúde Pública*, v.23, n.12, p.3041-3050. 2007.

COILE, Russell C.
Lotka's frequency distribution of scientific productivity. *Journal of the American Society for Information Science*, v.28, n.6, p.366-370. 1977.

CONSTANS, Carine; RIVOAL, Isabelle.
Le prestigie des revues scientifiques et la logiques de classement. In: Hurlet, Frédéric; Rivoal, Isabelle; Sidéra, Isabelle. *Le prestige*: autour des forms de la différenciation sociale. Paris: Boccard. p. 255-270. 2014.

COSTAS, Rodrigo; BORDONS, María.
The h-index: advantages, limitations and its relation with other bibliometric indicators at the micro level. *Journal of Informetrics*, v.1, n.3, p.193-203. 2007.

COURTIAL, Jean-Pierre; LAW, John.
A co-word study of artifical intelligence. *Social Studies of Science*, v.19, n.2, p.301-311. 1989.

COZZENS, Susan.
Taking the measure of science: a review of citation theories. *Newsletter of the International Society for the Sociology of Knowledge*, v.7, n.1, p.16-21. 1981.

CRONIN, Blaise.
Scholars and scripts, spoors and scores. In: Cronin, Blaise; Sugimoto, Cassidy. *Beyond bibliometrics*: harnessing multidimensional indicators of scholarly impact. Cambridge: The MIT Press. p.3-21. 2014.

CRONIN, Blaise; SUGIMOTO, Cassidy.
Beyond bibliometrics: harnessing multidimensional indicators of scholarly impact. Cambridge: The MIT Press. 2014.

DE BELLIS, Nicola.
History and evolution of (biblio)metrics. In: Cronin, Blaise; Sugimoto, Cassidy. *Beyond bibliometrics*: harnessing multidimensional indicators of scholarly impact. Cambridge: The MIT Press. p.23-44. 2014.

DESROSIÈRES, Alain.
The politics of large numbers: a history of statistical reasoning. Cambridge: Harvard University Press. 1998.

DONAVAN, Claire.
The qualitative future of research evaluation. *Science and Public Policy*, v.34, n.8, p.585-597. 2007.

EDITORIAL.
Editorial. *Medical History*, v.53, n.1, p.1-4. 2009.

EGGHE, Leo.
Theory and practice of the g-index. *Scientometrics*, v.69, n.1, p.131-152. 2006.

ELKANA, Yehuda et al. (Ed.).
Toward a metric of science: the advent of science indicators. Brisbane: John Wiley. 1978.

FERREIRA, Renata C.; ANTONELI, Fernando; BRIONES, Marcelo R.S.
The hidden factors in impact factors: a perspective from Brazilian science. *Frontiers Genetics*, v.4, n.130, p.1-2. 2013.

FRIGERI, Mônica. MONTEIRO, Marko Synésio Alves.
Qualis Periódicos: indicador da política científica no Brasil? *Estudos de Sociologia*, v.19, n.37, p. 299-315. 2014.

FURNER, Jonathan.
The ethics of evaluative bibliometrics. In: Cronin, Blaise; Sugimoto, Cassidy. *Beyond bibliometrics*: harnessing multidimensional indicators of scholarly impact. Cambridge: The MIT Press. p.85-108. 2014.

GARFIELD, Eugene.
The evolution of the Science Citation Index. *International Microbiology*, v.10, n.1, p.65-69. 2007.

GARFIELD, Eugene.
Is citation analysis a legitimate evaluation tool? *Scientometrics*, v.1, n.4, p.359-375. 1979.

GARFIELD, Eugene.
Citation analysis as a tool in journal evaluation. *Science*, v.178, n.4060, p.471-479. 1972.

GARFIELD, Eugene.
Citation indexes for science: a new dimension in documentation through association of ideas. *Science*, v.122, n.3159, p.108-111. 1955.

GINGRAS, Yves.
Les derives de l'évaluation de la recherche: du bon usage de la bibliométrie. Paris: Raisons d'Agir. 2014.

GINGRAS, Yves.
Les systèmes d'évaluation de la recherche. *Sciences de l'Information*, v.46, n.4, p.34-35. 2009.

GLÄNZEL, Wolfgang.
Seven myths in bibliometrics about facts and fiction in quantitative science studies. *Issi Newsletter*, v.4, n.2, p.24-32. 2008.

GODIN, Benoît.
The making of science, technology and innovation policy: conceptual frameworks as narratives, 1945-2005. Quebec: Centre Urbanisation Culture Société. 2009.

GRUPO SCIMAGO.
El índice h de Hirsch: aportaciones a un debate. *El Profesional de la Información*, v.15, n.4, p.304-306. 2006.

HAGSTRON, Warren O.
The scientific community. New York: Basic Books. 1965.

HARZING, Anne-Wil.
The Publish or Perish book. Melbourbe: Tama Software Research. 2011.

HE, Qin.
Knowledge discovery through co-word analysis. *Library Trends*, v.48, n.1, p.133-159. 1999.

HICKS, Diana et al.
The Leiden Manifesto for research metrics. *Nature*, v.520, p.429-431. 2015.

HIRSCH, Jorge E.
A index to quantify an individual's scientific research output. *Proceedings of the National Academy of Science*, v.102, n.46, p.16569-16572. 2005.

HUANG, Mu-hsuan; CHI, Pei-shan Chi.
A comparative analysis of the application of h-index, g-index, and a-index in institutional-level research evaluation. *Journal of Library and Information Studies*, v.8, n.2, p.1-10. 2008.

KUHN, Thomas.
A estrutura das revoluções científicas. São Paulo: Perspectiva. 1989.

LETA, Jacqueline.
Indicadores de desempenho, ciência brasileira e a cobertura das bases informacionais. *Revista USP*, n.89, p.62-67. 2011.

LEYDESDORFF, Loet.
The challenge of scientometric: the development, measurement, and self-organization of scientific communications. New York: Universal. 2001.

MANCHIKANTI, Laxmaiah et al.
Medical journal peer review: process and bias. *Pain Physician*, v.18, n.1, p.E1-E14. 2015.

MARTIN, Ben R.; IRVINE, John.
Assessing basic research: some partial indicators of scientific progress in radio astronomy. *Research Policy*, v.12, n.2, p.61-90. 1983.

MATTEDI, Marcos A.
Sociologia e conhecimento: introdução à abordagem sociológica do problema do conhecimento. Chapecó: Argos. 2006.

MATTEDI, Marcos A.; SPIESS, Maiko R.
Modalidades de regulação da atividade científica: uma comparação entre as interpretações normativa, cognitiva e transacional dos processos de integração da comunidade científica. *Educação e Sociedade*, v.31, n.110, p.73-92. 2010.

MCCAIN, Katherine W.
Mapping authors in intellectual space: a technical overview. *Journal of the American Society for Information Science*, v.41, n.6, p.433-443. 1990.

MCROBERTS, Michael H.; MCROBERTS, Barbara R.
A re-evaluation of Lotka's Law of scientific productivity. *Social Studies of Science*, v.12, n.3, p.443-450. 1982.

MERTON, Robert K.
The Matthew Effect in science. *Science*, v.159, n.3810, p.56-63. 1968.

MOWERY, David C.; ROSENBERG, Nathan.
Trajetórias da inovação. Campinas: Unicamp. 2005.

NARIN, Francis.
Evaluative bibliometrics: the use of publication and citation analysis in the evaluation of scientific activity. Washington: National Science Foundation. 1976.

OCDE.
Organização para a Cooperação e Desenvolvimento Económico. *Manual di Frascati*. Coimbra: Gráfica de Coimbra. 2007.

OKRASA, Wlodzimierz.
Differences in scientific productivity of research units: measurement and analysis of output inequality. *Scientometrics*, v.1, n.3, p.221-239. 1987.

PENDLEBURY, David A.; ADAMS, Jonathan.
Comments on a critique of the Thomson Reuters journal impact factor. *Scientometrics*, v.92, n.2, p.395-401. 2012.

PINHEIRO, Lena V.R.
Lei de Bradford: uma reformulação conceitual. *Ciência da Informação*, v.12, n.2, p.59-80. 1983.

POPPER, Karl R.
A lógica da pesquisa científica. São Paulo: Cultrix. 2006.

PORTER, Theodore M.
Trust in numbers: the pursuit of objectivity in science and public life. New Jersey: Princeton University Press. 1995.

PRICE, Derek de Solla.
Toward a model for science indicators. In: Elkana, Yehuda et al. (Ed.). *Toward a metric of science*: the advent of science indicators. Brisbane: John Wiley. p.69-95. 1978a.

PRICE, Derek de Solla.
Editorial statements. *Scientometrics*, v.1, n.1, p.3-8. 1978b.

PRICE, Derek de Solla.
Little science, big science. New York: Columbia University Press. 1963.

SAISANA, Michaela; D'HOMBRES, Beatrice.
Higher education rankings: robustness issues and critical assesment. Luxembourg: Office for Official Publications of the European Communities. (JRC Scientific and Technical Reports). 2008.

SCHREIBER, Michael.
An empirical investigation of the g-index for 26 physicists in comparison with the h-index, the 4-index and the r-index. *Journal of the American Society for Information Science and Technology*, v.59, n.9, p.1513-1522. 2008.

SELEK, Salih; SALEH, Ayman.
Use of h-index and g-index for American academic psychiatry. *Scientometrics*, v.99, n.2, p.541-548. 2014.

SHATZ, David.
Peer review: a critical inquiry. New York: Rowman and Littlefield. 2004.

SGUISSARDI, Valdemar; SILVA JÚNIOR, João dos Reis.
Trabalho intensificado nas federais: pós-graduação e produtivismo acadêmico. São Paulo: Xamã. 2009.

SILVA, Antonio O.
A sua revista tem Qualis? *Mediações*, v.14, n.1, p.117-124. 2009.

SMITH, Richard.
Peer review: a flawed process at the heart of science and journals. *Journal of the Royal Society of Medicine*, v.99, n.4, p.178-182. 2006.

THOMSON REUTERS.
2015 Journal Citation Reports. Disponível em: http://wokinfo.com/products_tools/analytical/jcr/. Acesso em: 26 out. 2015. 2015.

TOL, Richard S.J.
A rational, successive g-index applied to economics departments in Ireland. *Journal of Informetrics*, v.2, n.2, p.149-155. 2008.

URBIZAGÁSTEGUI ALVARADO, Rubén.
A Lei de Lotka na bibliometria brasileira. *Ciência da Informação*, v.31, n.2, p.14-20. 2002.

USHER, Alex; SAVINO, Massimo.
A global survey of university ranking and league tables. *Higher Education in Europe*, v.32, n.1, p.5-15. 2007.

VAN RAAN, Anthony F.J.
Measuring science: capita selecta of current main issues. In: Moed, Henk F.; Glänzel, Wolfgang; Schmoch, Ulrich. *Handbook of quantitative science and technology research*: the use of publication and patent statistics in studies of S&T systems. Dordrecht: Kluwer. 2004.

VANTI, Nadia A.P.
Da bibliometria à webometria: uma exploração conceitual dos mecanismos utilizados para medir o registro da informação e a difusão do conhecimento. *Ciência da Informação*, v.31, n.2, p.152-162. 2002.

VELHO, Lea M.S.
Como medir ciência? *Revista Brasileira de Tecnologia*, v.16, n.1, p.35-41. 1985.

VINKLE, Peter.
The evaluation of research by scientometric indicators. Oxford: Chandos. 2010.

WHITLEY, Richard.
The intellectual and social organization of the sciences. Oxford: Oxford University Press. 2000.

WHITLEY, Richard.
The establishment and structure of science as reputational organization. In: Elias, Norbert; Martins, Hermínio; Whitley, Richard. *Scientific establishments and hierarchies*. London: D. Reidel. p.313-358. 1982.

WOUTERS, Paul.
The citation culture. Tese (doutorado em Estudos da Ciência e Tecnologia) – University of Amsterdam, Amsterdam. 1999.