

Academic Evaluation: Universal Instrument? Tool for Development?

Mariela Bianco¹ · Natalia Gras^{1,2} · Judith Sutz¹

Published online: 8 August 2016
© Springer Science+Business Media Dordrecht 2016

Abstract Research agendas and academic evaluation are inevitably linked. By means of economic incentives, promotion, research funding, and reputation academic evaluation is a powerful influence on the production of knowledge; moreover, it is often conceived as a universal instrument without consideration of the context in which it is applied. Evaluation systems are social constructions in dispute, being the current focus of international debates regarding criteria, indicators, and their associated methods. A universalist type of productivity indicators is gaining centrality in academic evaluation with profound effects on the content of research that is conducted everywhere. Specifically, evaluation systems based on this type of indicators are sending negative signals to scientists willing to conduct research on contextualized agendas, particularly those negotiated with non scientists. On the basis of theoretical and empirical studies documented on the specialized literature and extensive personal engagement with university research policy in Uruguay, we argue that the consolidation of evaluation practices of alleged universal validity deteriorates and discourages a type of research which is undeniably important in developing contexts.

Keywords Academic evaluation · Research agendas · Research for development · Science in developing countries · Uruguay

✉ Mariela Bianco
mbianbo@gmail.com

¹ Academic Unit of the University Research Council (CSIC), Universidad de la República, Jackson 1303, 11200 Montevideo, Uruguay

² Autonomous Metropolitan University (UAM-Xochimilco), Mexico City, Mexico

Introduction

Academic research has become an international endeavor in which collaboration across the globe is an extended practice. Collaboration, though, occurs in an uneven playing field. Asymmetries among regions and countries are obvious in terms of research resources (funds, infrastructure, scientists) concomitantly leading to asymmetries of power in setting research agendas. These are heavily influenced by the challenges posed at research systems of hegemonic centers and regions and expanded everywhere through the logic of global collaboration. Further, a universalist type of academic evaluation based on productivity indicators has become a worldwide trait of the science system. Wherever researchers work, their academic and material rewards are increasingly based on it.

Accordingly, contextualized research agendas dealing with problems of local relevance but of little interest for mainstream science may face difficulties getting established. In turn, stubborn researchers who pursue them are left in a difficult academic situation by evaluation systems. A contention of this paper is that this situation jeopardizes the possibilities of using research results for development purposes. This assertion is in line with the answer given to the following question: what is the influence of the type of knowledge produced, if any, on the probability that research results and capacities are put into use in development processes? If the answer is that it makes no difference and therefore any type of knowledge is the same, no matter what questions or problems are included in the research agenda, the topic of this paper may be useless. However, if the type of academic research effectively conducted is considered important in order to assess the contribution of knowledge to development, our initial contention holds.

Understanding how research inquiries are configured and how research agendas are set has always been part of the core of the history, sociology and economics of academic research. There are several and strong discrepancies regarding the main influences that shape research agendas. Some scholars have focused on the dominant interests in particular socio-historical contexts (Hessen 1931; Bernal 1994) while others have examined institutional dynamics associated with academic research and the puzzles that science formulates (Merton 1942; Kuhn 1970). Still other authors indicate that research influences are intertwined in a complex set of internal and external factors in relation to science (Rosenberg 1982; Bunders 1987; Sábato and Botana 1968).

The purpose of this paper is to analyze one particular set of influences on research agendas: academic evaluation systems. We are interested in the way evaluation systems and practices are set and how they influence research agendas. Evaluation systems are social constructions in dispute, being the current focus of international debates regarding criteria, indicators and their associated methods of measurement; their consequences on the academic enterprise are equally analyzed and disputed.¹

¹ According to Barré (2010), indicators are not a reality or a technical and scientific truth. Their construction results from a political process because they are based on a specific model (among many others) in order to understand how science works or how it should work. Simultaneously, interpreting these indicators requires multiple judgments where experts involved also introduce their values, rules and

We are aware that research evaluation is only one among several influences on research agendas. However, the direct linkage between research agendas and academic evaluation is strong enough to make the analysis relevant from a development perspective. In the words of Hess (2007: 29, emphasis added): “In my view the primary question for science and technology studies in an era of globalization is no longer the constructivist question of how scientific knowledge is socially negotiated or shaped, but instead *the structural question of what science is selected to be done. To begin to answer that question, we must first turn to the reward system in science.*”

A problematic aspect of this topic is that evaluation is usually presented as valid for scientific research in general, regardless of the context in which it is conducted. Following Hess, an evaluation system understood as a universal tool and another one conceived as a development tool may give rise to partially different research agendas. In this article, we explore this topic drawing on an extensive review of specialized literature, analyzing the case of Uruguay² and reflecting upon personal engagement of the authors with university research policy. In Sect. 2 we elaborate on research evaluation methods and current associated debates. In Sect. 3 we analyze the Uruguayan research environment and several tensions that are likely to appear when the implementation of contextualized research policies, more or less tuned with development goals, are coupled with academic evaluation strongly based on conventional metric indicators. The final section offers some preliminary remarks and follow-up questions regarding research evaluation for development.

Academic Evaluation and the Production of Knowledge: A Glimpse at the State of the Discussion

General Aspects

International literature on the effects of evaluation systems on the production of knowledge is profuse, focusing on science general aspects as well as on specific disciplines (Whitley 2007; Martin and Whitley 2010; Gläser and Laudel 2007a, b; van der Most 2010; de Jong et al. 2011; Regeer et al. 2009; Hemlin and Barlebo Rasmussen 2006; Bunders and Broerse 1991; Bunders 1990; Elzinga 1988; Sahel 2011; Hicks 2004, 2006, 2013; Katz 1999; van Dalen and Henkens 2012, among others). Nonetheless, little is still known about the way evaluation systems affect the contents of research. This article does not fill this cognitive gap but attempts to contribute in the general direction of what type of science is favored or discouraged by prevailing evaluation systems. By the time we were finalizing this article, Gläser

Footnote 1 continued

personal visions to make decisions. In this sense, evaluation as well as the development and interpretation of indicators relate to a normative and, therefore, debatable process.

² Uruguay is a high-income country in terms of GDP per inhabitant according to the World Bank classification. However, it belongs to what can be considered the developing world in terms of STI issues.

and Laudel (2016) published a review highlighting what is still not fully understood about the relation between science governance and changes in research contents.

A useful analytical scheme, proposed by Whitley (2007), considers evaluation effects on the production of knowledge taking into account the characteristics of the context. This is summarized in Table 1.

Table 1 facilitates the understanding of the interrelated character of contextual elements. Let us suppose a strong research evaluation system - with high frequency of performance evaluations, with highly formalized rules and procedures, and whose results are public. Let us also imagine that the public S&T system is organized around competitive project-based funding and that research fields have a few available funding sources and legitimating audiences. Combined with other contextual elements, this strong research evaluation system will probably reinforce the segmentation of research organizations (Whitley 2007) resulting in what is known as the Matthew effect on science³. Other results will likely be an intensified competition to gain recognition, an increased importance of central disciplines, their standards and research priorities, a decline in intellectual diversity and cognitive pluralism and, accordingly, a plausible trend towards inhibiting the development of new fields and approaches. A similar outcome may be obtained if the scientific elite is strong and cohesive, playing an important role in the implementation of evaluation systems. The opposite may occur if in the public S&T system and funding, programs oriented by public policy objectives are significant, there is a variety of funding agencies and goals, and academic research organizations have strategic autonomy and control over their resources.

The problem is even more complex as the characteristics of the public science system, the scientific fields as well as the research evaluation system are all influenced by the distribution of authority relations (Whitley 2010) among the diverse actors in the STI system (researchers, scientific elites, public policy-makers, research funders, users). Authority relations are a direct influence on the selection of research objectives, shaping at the same time the influences of evaluation systems in a particular time and place.

A lot has been written about the way a particular set of changes implemented after World War II altered the dynamics and logics of knowledge production, especially in highly industrialized countries. Science increasing costs were a major cause of changes resulting in a relative decrease in research public funding and an increase in the competition for funds as well as greater steering of funds to specific goals (Ziman 1994; Whitley 2010). These transformations have reconfigured authority relations changing the context in which researchers make their strategic decisions regarding research topic selection, goals, methods, as well as their participation in research networks with colleagues (national or international) or with non-academic actors (civil society, firms, government), and the selection of communication channels for research results (Gläser and Laudel 2007b). In general

³ Merton (1988) coined this expression in direct reference to a passage from the Gospel of Matthew: "Whoever has will be given more, and they will have abundance. Whoever does not have, even what they have will be taken from them." In science, that means that those who have strong scientific capacities will be given more opportunities to strengthening them further as a result of the academic system structure, while those whose capacities are weak will be left with even less.

Table 1 Contextual elements to be considered in the analysis of research evaluation impacts

Contextual elements	
Characteristics of public S&T system	<p>Research funding regime (competitive funds, national long term thematic program funding, others)</p> <p>Strategic capacities and research organization autonomy for decision making</p> <p>Degree of segmentation of research organizations and of researchers labor market</p>
Characteristics of the scientific fields	<p>Diversity of available funding agencies</p> <p>Diversity of audiences that legitimize research results</p> <p>Centrality and prestige of the field in comparison to other scientific fields</p> <p>Cohesion and prestige of scientific elites of the field</p>
Central characteristics of predominant research evaluation system	<p>Governance and structure of the system (frequency of performance evaluation, standardization of criteria, formalization of procedures, evaluation units: programs, projects, teams, researchers)</p> <p>Linkage of evaluation results with research funds allocation</p> <p>Communication of evaluation results (open, partially restricted, restricted)</p>

Source: adapted from Whitley (2007)

terms, there has been a loss of relative autonomy of academic institutions and researchers regarding their strategic decisions associated with changes in the governance of research institutions, particularly in universities (Whitley 2010; Dobbins and Knill 2014). Although the above changes have been mostly identified and analyzed for the developed world, they are equally present in underdeveloped contexts (Bensusán et al. 2014).

Constitutive Effects

The quality of research results is a basic criterion in academic research evaluation. However, research quality is not a one-dimensional concept. Among multiple authors who have addressed this issue, we select the characterization of Gläser and Laudel (2007b) regarding the epistemic properties of research. These authors propose to consider the following properties of research: (1) type of research (for example, methodological, theoretical, experimental or field research); (2) relationship to the community's majority opinion (non-conformist versus mainstream); (3) time characteristics of research (long-term versus short-term processes); (4) degree of heterogeneity of knowledge combined in research (interdisciplinary); (5) degree of intellectual risk taken in the research; (6) reliability of results.

Not all evaluation systems consider this set of properties or weight them equally. The signals from evaluation systems regarding the "quality expectations" are likely to induce changes in research strategies as researchers adapt to them. "The adaptation of research strategies and approaches to the 'quality expectation' is likely to change more than research 'quality' as measured by the system" (Gläser and

Laudel 2007b: 130). Observed effects that have become *constitutive effects* (Dahler-Larsen 2014) of evaluation systems on epistemic characteristics of research and other aspects of knowledge production have proliferated in the specialized literature (Whitley 2007; Hicks 2004; Martin and Whitley 2010; Elzinga 1988; van der Most 2010; Bianco et al. 2014a). The British Research Assessment Exercise (RAE), established in the mid-1980s, offers a paradigmatic case to study *constitutive effects*. Table 2 summarizes some of these effects documented after two decades of implementation of the RAE.

The Distortion of the Methods

Diverse authors agree that stimulating diversity and quality in all cognitive fields requires different forms of evaluation of research results that take into account epistemic differences and research goals, rewarding results accordingly (Bianco

Table 2 Insights on the systematic application of a “strong” evaluation system

1. *Effects show wide variation across fields* Disciplines well suited to perform according to the RAE include: (a) disciplines in which quality was already closely linked to a hierarchy of journals (such as laboratory based and economics), (b) those where there was standardization of technical entry requirements, (c) with more propensity to engage in incremental research (‘normal science’ linked to an established paradigm), (d) with a strong refereeing culture, and often a weaker relationship between research and teaching. In contrast, for many humanities and social sciences, characterized by more integration between research and teaching and where research may yield a variety of outputs other than academic publications, the cultural and structural consequences were more traumatic
2. *Changes in research direction and characteristics* Individuals are more constrained in pursuing their own research agendas. Research has experienced notables changes; in general researchers are induced to focus on: (a) shorter term rather than longer term research, (b) incremental rather than more ambitious or open-ended pioneering research, (c) mainstream rather than alternative research or research in a highly specialized sub- field, (d) monodisciplinary rather than multi or interdisciplinary research, (e) academic rather than professional research such as in medicine or law, (f) research that yields journal articles rather than books, (g) research where the results can be published in top journals rather than more specialist and/or lower status ones
3. *Distortions derived from the pressure to publish* High frequency of evaluations and growing publication pressures have reinforced a tendency to premature publication before the research is fully ready. Those with poor publishing records, according with predominant criteria, suffered strong individual consequences
4. *Degradation of the role of teaching* RAE incentives are perceived to be much stronger than those for teaching with deleterious effects on teaching activities. Leading researchers tend to devote less time to lecture preparation, meeting with students or preparing new courses. Some have even managed to negotiate research only contracts with their universities
5. *Degradation of involvement with other academic activities (different from research)* There is less volunteering by faculty to participate in different university committees, administration, community work and outreach, writing popular books, providing policy advice, among others, with a general decrease in organizational loyalty
6. *Impact on private life* RAE’s pressure has encouraged overwork and increasing levels of stress. RAE has disadvantaged researchers (predominantly women) who took time off work for family or personal reasons, resulting in a gap in their published output. Colleagues are less willing to take over the work load left by those who are forced by diverse circumstances to work part time because this may jeopardize their chances to meet the required publication level

Source: adapted from Martin and Whitley (2010)

et al. 2014a; Hicks 2013; de Jong et al. 2011; Regeer et al. 2009; Hemlin and Barlebo Rasmussen 2006; Bunders and Broerse 1991; Bunders 1990; Elzinga 1988). Nevertheless, qualitative peer review and evaluation based on bibliometric methods are the dominant forms of research evaluation worldwide. Also, individual academic performance evaluation increasingly rests on hybrid evaluation practices: judgments elaborated by peers are commonly informed by the revision of academic merits on the basis of bibliometric indicators.

These forms prevail despite the fact that alternatives are available. For instance, evaluation in the context of application of results, known as *quality monitoring* (Hemlin and Barlebo Rasmussen 2006), is an alternative form of evaluation to peer review and to metrics-based *-quality control-*, even if it is not a widespread practice. Knowledge legitimacy is not exclusively associated with internal scientific value (originality and methodological rigor) but it also rests on external factors such as its contribution to industrial development, public policy design, among others. According to these authors, the focus of research evaluation needs to shift from individual researchers to organizations and networks, to induce socially robust knowledge assessed by a variety of evaluators combining experts, users and lay persons. However, the use of bibliometric indicators has steadily increased and turned evaluation into a practice now led by the data rather than by judgments (Hicks et al. 2015).

Ironically, metrics' centrality in academic evaluation does not result from a consensus about its validity. On the contrary, several voices have been raised against its pertinence as a unique method. For example, in 2012, during the annual meeting of the American Society for Cell Biology, a group of scientists and journal editors wrote the *Declaration on Research Assessment: Putting science into the assessment of research* (DORA 2012), which was later signed by several research communities, university authorities, policy-makers, and science journal editors across the world.⁴ DORA recommends not to use journal-based metrics, such as Journal Impact Factors "to assess an individual scientist's contributions, or in hiring, promotion, or funding decisions." Previously, a research report warned on the use of bibliometrics as the unique simple method for evaluation, stating the following. "There is a risk that any metrics exercise may be intrinsically self-defeating, because it depends on indicators as proxies for the activity of interest. Once an indicator is made a target of policy, it starts to lose the information content that originally qualified it to play such a role. There is room for manipulation, there may be emergent behavioral effects and the metrics only capture part of the research process and its benefits" (Evidence, LTD. 2007: 35, quoted by McNay 2009: 49). More recently, *Nature* has published in its Comment section the Leiden Manifesto for research metrics, which reports the pervasive misapplication of bibliometric indicators to the evaluation of scientific performance and develops ten principles for the best practice in metrics-based research assessment (Hicks et al. 2015).

One wonders why, despite so much criticism, bibliometric methods prevail in research evaluation worldwide. Gläser and Laudel (2007a) indicate that over the last

⁴ In May 2013, the *Science* editor quoted the San Francisco Declaration and emphasized the need to stop using impact factor in research assessments (Alberts 2013).

two decades, the diffusion of bibliometric evaluations has been driven by the adoption of the paradigm of the new public management by science policy. This paradigm centers on the belief that market competition and market exchange are the best ways of conducting any public task regardless of its content, and that there is no better way of solving allocation problems or producing efficiently. The premises of the new public management: efficiency, transparency, accountability, quality and competitiveness, dominated the public sector management. Accordingly, public science funds can be managed and allocated in the same way than those of any other public policy driving several research organizations to mimic what the private sector purports as best practices (Whitley 2010). The same situation occurred in England, Mexico, Switzerland and Germany (Martin and Whitley 2010; Morris 2010; Bensúsán et al. 2014; Benninghoff and Braun 2010; Meier and Schimank 2010; Schimank 2005).

Gläser and Laudel (2007a) outline three main reasons for the rapid increase of quantitative indicators of research performance:

1. The increasing demand for everyday review activities for competitive funding and publications makes it unfeasible for peer review alone to meet the demand for evaluation. Bibliometric indicators are perceived as *cheaper* and *faster* than peer review.
2. Bibliometric evaluations appear to be legitimated by scientific practice; thus they are perceived as more *trustworthy* and *objective* than peer review (they seem to control bias derived from personal interests and authority relations in evaluation).
3. Bibliometric evaluations give the impression of being *accessible* by politicians and managers without the involvement of scientists. In peer review, recommendations are inextricably linked to practices and characteristics of the science fields. Instead, the numbers produced by quantitative evaluations appear to be decontextualized and for the same reason can be more easily processed by science managers than qualitative judgments.

These perceptions about the convenience, accessibility and objectivity of quantitative indicators, ignore the fact that the results of metrics need to be interpreted by scientists of the field and cannot be equally used at all levels of aggregation. Gläser and Laudel (2007a) elaborate on the missing modalities of bibliometric methods. Modalities refer to the conditions under which the methods can be applied and the way in which they must be used in order to produce valid and reliable results. If some of these conditions are missing, the reliability and validity of data deteriorate. These authors analyze five conditions that are often missed when bibliometric indicators are used in research evaluations: (a) citation-based methods do not measure quality *per se* but rather one proxy to an important aspect of quality; (b) they are valid only if they statistically represent impact; (c) in order to draw valid conclusions about research performance, the whole research output of the evaluated unit must be covered by the assessment (one missing well-cited publication can create a significant error); (d) since publications reach their highest citation rate after at least three years, the validity and reliability of evaluation methods is highly dependent on the time-frame chosen for analysis; (e) bibliometric

measures depend upon the specificity of knowledge production in different fields, therefore cannot be “compared nor aggregated without normalizing the results with field-specific reference values” (Van Raan 1996: 403, quoted in Gläser and Laudel (2007a)).

In turn, qualitative peer review has been often criticized for its subjectivity and put into question on the basis of plausible distortions associated to the corporate behavior of scientific elites. Moreover, peer review faces a serious time constraint in a context of increasing demand from simultaneous evaluation processes and, eventually, an additional problem of expert recruitment because reviewing is time consuming and finding the right expertise is not always easy. Time issues can have an effect on the quality of research evaluation, for example, inducing superficial reviews of scientific contributions (Sahel 2011). These problematic aspects of peer review have a worldwide scope and have been detected in developed and developing countries.

Metrics have intrinsic restrictions and correctly applying quantitative indicators involves conditions that often undermine its alleged economy, reliability and accessibility. Peer review is well suited for field specificity and the holistic character of academic output, but faces credibility and feasibility issues. In the context of the “audit society” (Power 1987) shortcomings are often overlooked in the name of the urgent need for selecting, assessing, rewarding and punishing. Indeed, performance evaluation pressures lead to an intertwining of metrics and peer reviews in everyday evaluation practices. It is not uncommon that research evaluation committees end up relying on publication counts and available indexes calculated on them instead of producing their own qualitative assessments of research results. In this sense, researchers’ hiring and promotion increasingly depend upon bibliometrics used by peers while qualitative assessments are growingly restricted to manuscript evaluation in journal review boards and research proposal evaluation for competitive funds. Acknowledging this trend does not deny the fact that research evaluation is a fundamental part of the academic enterprise. On the contrary, it is precisely for this reason that a critical appraisal of current research evaluation practices is needed as part of a search for alternatives that is now urgent. We turn now to such appraisal in the academic environment of a developing country, Uruguay.

When Universalist Criteria Clashes with the Context: Reflections from the Uruguayan Case

The literature review just presented shows increasing uneasiness and concerns about the consequences of academic reward systems. While this literature mainly examines the situation of highly industrialized countries, several of the concerns it raises are quite universal. For instance, the features described in the assessment of the British RAE related to the increasing stress of academic life pushed by the paper count bias of the evaluation system, is as vivid there as it is in most universities in developing countries. Strategies for academic survival are easily recognized everywhere: partitioning papers, not fully backed submissions to accelerate the rate

of publications, multiple authorships, among others. The reasons put forward to rely more heavily on quantitative measures of research quality are also quite expanded; moreover, they may be even more strongly supported in weak research systems. The smallness of a research community makes peer review suspicious of lack of independent criteria; when a community, as it often occurs, writes proposals and publishes in languages different from English, the universe of peer reviewers diminishes notoriously, adding to the dismissive arguments. In addition, the smaller the community of peers, the heavier the time burden of qualitative evaluation. Thus, notwithstanding the fact that specific working conditions of researchers vary markedly among countries and among academic settings within countries, the main features and trends affecting researchers' academic lives may be found everywhere. However, when we move from individual researchers to more aggregate levels, for instance, the setting of research agendas, the similarities weaken. There are academic milieus that, either because they have more funding, better infrastructures or accumulated prestige, are remarkably powerful to set research agendas. The themes, methodological approaches, disciplines, research questions around which the mainstream research agendas get established are mainly put forward in those academic milieus. A vast and diverse "academic periphery," characterized by a lower academic strength of the research premises and/or by lateral themes and approaches that researchers choose to follow, deploys outside such central hub. This academic periphery also includes institutions and researchers working within the central hub; however, a more systemic situation can be seen in developing countries. There, a vast majority of academic institutions belong to such periphery, in part due to their relative weakness, measured through several indicators, and in part because they focus on relevant topics for the local context which are relatively uninteresting for mainstream research. Vessuri et al. (2014) alert that in Latin America and other peripheral regions, the expansion of citation-based indicators generates a competition regime in which quality is linked to journal rankings owned by private publishers. They suggest that the use of these indicators in career evaluation procedures works against scientific creativity and originality, and may even jeopardize the general quality of research as well as its utility for development. Thus, in developing countries in particular, the clash between universalist evaluation criteria and contextualized research policies produces tensions that may negatively affect the success of the latter. A universalist type of evaluation equates high quality research with research producing articles for peer-reviewed indexed international journals with little attention paid to any other considerations.

Development processes require knowledge that results from contextualized research policies. When policies, already difficult to put forward due to the already mentioned weakness of research systems, are further undermined by universalist academic evaluation systems, development processes suffer. In this section we examine how this clash is manifested in a small developing country like Uruguay.

The Uruguayan Research System: A Synthetic Approach

Three features of the Uruguayan research system are particularly worth stressing. The first is that about 80% of all research is conducted at the Universidad de la República (UdelaR) (Dicyt-MEC 2012), which until 2014 was the only public university in the country. This university, founded in 1849, had a marked professional orientation. Academic research in natural and exact sciences as well as in social sciences found a small place in the classical faculties of Law, Medicine, Engineering, Agriculture and Veterinary Sciences until twenty years ago. Partly as a consequence, academic postgraduate programs were hardly developed until the 1990s, implying that a high proportion of researchers followed their postgraduate studies abroad. It is important to note the resilience of this first feature: during the last 30 years many new institutions where research is performed were created, including some private universities; however, the overwhelming majority of research continues to be conducted at UdelaR. It is worth highlighting too that UdelaR is the only research institution in Uruguay where all areas of knowledge are cultivated; it is autonomous to allocate its budget and to set its research priorities.

A second feature worth stressing relates to a specific trait of university research: all staff at UdelaR is hired through the assessment of applications presented to open calls. Once in the faculty, professors can apply to the Full-Time Program that evaluates applicants on the basis of a plan in which research is a major component. The Full-Time program offers an important economic incentive (60% salary increase) in reward for an integral dedication to university work, including teaching, research and outreach. Since its creation in the late 1950s, the Full-Time Program fosters knowledge production and is renewed fundamentally based on the qualitative evaluation of the research results over periods of five years. Up to the creation of the National System of Researchers (NSR) in 2008, the Full-Time Program provided the main set of criteria to evaluate individual academic research performance.

The third feature refers to the utilization of Uruguayan research capacities, both by industry and by government: it is quite weak. This fact relates, in the case of industry, to the country's productive structure, mainly based on imported technologies. Except in some agricultural fields, little efforts are devoted to the development of indigenous and eventually better adapted knowledge solutions. In the case of government, a structure characterized by fragmented autonomies fosters the isolated acquisition of ready-made solutions, weakening the participation of local research capabilities. These features reinforce the syndrome of "university loneliness" (Arocena and Sutz 2010), meaning that research agendas do not usually receive clearly formulated demands from outside the cloisters. As in most developing countries, the fact that Uruguay is not a "knowledge-based and innovation-driven economy" (de la Mothe and Paquet 1996) is a contextual feature with important consequences for knowledge production and use.

Some of the features described are illustrated in the S&T figures for Uruguay and some other countries summarized in Table 3. Overall, Uruguay has a low R&D expenditure, a small research community, and a high concentration of researchers

Table 3 S&T figures for Uruguay and selected countries

	R&D/GDP (%)			Researchers (FTE) per 1000 inhabitants in the economically active population			% of researchers (FTE) working in institutions of higher education	
	2000	2006	2012	2000	2006	2012	2000	2006
Uruguay	0.23	0.36	0.33	0.61	0.58 (2008)	1.08	90.0	–
Costa Rica	0.38	0.43	0.57	0.32 (2003)	0.58 (2008)	0.72	–	80.0 (2005)
Argentina	0.5	0.46	0.65	1.8	3.3	4.7	50.0	44.6
Spain	0.9	1.26	1.24	4.3	5.3	5.4	55.0	48.0
Portugal	0.72	0.95	1.37	3.2	4.4	7.7	51.3	48.8

Source: RICYT Indicators, www.ricyt.org

working in universities. We now turn to some major events that occurred during the last decades.

The Uruguayan research system has been rebuilt almost from ashes since the mid-1980s, after the catastrophic consequences of the military dictatorship of the period 1973–1984. A massive exodus of scientists followed political persecution and dismissal from university positions. The sinking of the research budget implied a decade of obsolescence for libraries and research facilities, making the academic situation at the eve of re-democratization extremely fragile. Since 1985 a strong institutional re-configuration took place with impetus: the almost destroyed research in natural and exact sciences was rebuilt, new university colleges were created, particularly the Faculty of Sciences and the Faculty of Social Sciences, graduate programs flourished, and new institutions devoted to research were established, among them the National Institute for Agricultural Research (INIA). Particularly significant was the effort devoted to the reconstruction of research and training in basic sciences. This was possible through the creation of the Program for the Development of Basic Sciences (PEDECIBA) in 1986, which pioneered the academic postgraduate studies in the country and put in place periodical evaluation exercises for researchers in the Program. Additionally, the small national budget devoted to R&D was fortified in the 1990s with loans from IDB; around the same time UdelaR established a University Research Council with specific competitive funds for research projects in all disciplines. Academic research revamped in the country at that time: one third of all the university research groups identified in 2001 were established between 1994 and 1997 (UA-CSIC 2003).

In 1998 a Researchers Fund was established, mainly to countervail the low salaries that academic researchers perceived (it had only two editions, in 1998 and 2004). A monetary supplement was awarded to those researchers considered the best among all researchers evaluated positively because the endowment of the Fund was relatively small, being able to contemplate a small portion of the latter. The establishment of this Fund was an opportunity for researchers in different disciplines to openly discuss what should count as academic merit. An interesting

debate confronted, in particular, the specificities of the different disciplines within the basic sciences. The number of citations was discussed, given that chemistry papers outweighed by far those in mathematics. The debate also involved differences between papers authored by one or two researchers, common in theoretical physics, versus “multitudinal” papers in which co-authors easily average more than ten, like in experimental biology. The variety of criteria required to assess applications to the Fund was acknowledged, but the criteria effectively used were mainly the number of papers published in academic journals, with special weight given to papers in English language.

In 2007 two important institutional innovations took place. The first was the creation of a National Agency for Research and Innovation (ANII), which regularized the competitive calls for research projects and implemented policy instruments aiming at fostering innovation. The second one was the establishment of a National System of Researchers (NSR), whose main difference with the former Fund was its comprehensive and permanent nature, meaning that (i) every researcher with sufficient academic merit, regardless of his/her institutional affiliation, would be integrated into the system and receive an economic bonus, and (ii) an evaluation procedure heavily based on research outputs was set up to periodically assess if researchers remain in the system, should be upgraded to higher levels or withdrawn from it.

The consolidation of the NSR can be seen as a watershed in the Uruguayan research system; in Whitley’s terms, a shift from a set of contextual elements to a different one. The different research settings comprising the core of the Uruguayan research system - the public University, the basic science development program, and the institute for agricultural research being the most conspicuous - enjoyed a fair degree of strategic capacity and research organization autonomy for decision making. The legitimizing spaces of research results were concomitantly diversified. Even if a selected group of researchers considered themselves an elite due to their publication records, there was no overall recognition of that. Researchers’ evaluation was relatively differentiated, either by research goals or orientation, given the important role played by institutional specific criteria. Before the NSR a relatively weak universalist research evaluation system prevailed at the country level. This was particularly evident at UdelaR, where researchers with quite different publication records - in number, type, and prestige of the publication source - integrated the same level of the Full-Time Program.

With the NSR a new scenario took shape that could be characterized as a “strong research system” in Whitley’s terms. A supra-institutional procedure to evaluate researchers was established, academically led by the above mentioned elite and guided by universalist criteria with an important quantitative bias. The results of the frequent evaluation rounds are made public and give rise to almost compulsive comparisons among researchers, a new trait in Uruguay’s research culture. Some funding decisions are tied to evaluation results: for instance, only researchers in the NSR are allowed to supervise graduate students on fellowships granted by ANII. It is too soon for any kind of impact assessment, but some indications suggest that strategies aimed at being accepted and further survive through the NSR are being

privileged above other concerns, with particular attention given to increasing the number of English language publications in indexed journals.

Following Whitley (2007), the higher frequency of evaluation of individual performance at NSR is expected to intensify competition for academic recognition and at the same time reinforce the preeminence of well established disciplines reducing cognitive diversity. Simultaneously, the low segmentation of research organizations in Uruguay combined with a relatively high autonomy for decision making in institutions are likely to favor cognitive pluralism in the research system. That is, the contextual elements of the Uruguayan system may intervene in opposite directions to strengthen or weaken the effects of the NSR. The final outcome between contradictory effects shall be seen in the future. So far, the Uruguayan institutional academic landscape reached a relatively stable stage: it is a good moment to analyze the kind of tensions that revolve around the academic evaluation system.

Academic Evaluation as a Source of Tensions and Discussions for Uruguayan Researchers

Research evaluation can be conceived as a system of signals that influences the research that gets conducted in a particular time and place. In this sense, Bianco et al. (2014a: 220) indicate:

An evaluation system not only develops after the fact judgments for research proposals competing for funds, articles submitted for publication, or final research reports. On a less direct way, but no less binding, an evaluation system sends signals to knowledge production influencing research orientation and agenda settings (Elzinga 1988; Whitley and Gläser 2007; van der Most 2010). In other words, the characteristics of the research evaluation process will have an impact, among others, on the organization and production of knowledge (Whitley 2007). In this sense, it can be asserted that evaluation steers research and thus, can stimulate, discourage, or at least not motivate some forms of knowledge production (own translation).

Thus, research communities will mostly orient their research according to the messages, more or less explicit, derived from research evaluation systems. Evaluation systems strongly based on universalist criteria may discourage research lines addressing contextualized problems dealing with development issues. These will most certainly be conducted in different languages (rather than in English), often require time frames that exceed the time available to report results for evaluation, and involve interaction processes with diverse non-academic agents. The association between evaluation traits and researchers' willingness to address local research topics and development related issues can be more clearly seen on the basis of a few tensions that we illustrate referring to the Uruguayan academic system.

As we have already stated, the Uruguayan academic community is rather small, implying that any rigorous academic evaluation process needs to involve foreign scholars. Moreover, a major proportion of senior researchers in almost all

disciplines has received graduate training abroad. These facts favored the development of deliberate strategies to strengthen the links between Uruguayan researchers and the world academic community as a way of creating research capabilities and enhancing research quality. Such strategies foster linkages with international research centers, encourage participation in research networks, and stimulate communication of research results in standardized scientific formats (international conferences, indexed journals). Understandably, all these efforts contribute to the simultaneous acquisition of research evaluation practices of alleged universal validity. As previously indicated, a “universalist” evaluation implies evaluating research regardless of the academic context in which it is conducted and relying heavily on metric-based indicators. As a complement, universalist evaluation may also indicate evaluating without consideration of disciplinary cultures with preeminence of dominant ways of communicating research results. For instance, assessing the practice of writing books which is characteristic in some academic disciplines with the criteria of other disciplines in which the best tradition is the production of journal articles.

Different tensions arise from the expansion of universalist evaluations. A first one emerges when Uruguayan researchers’ performance is measured against the standards of those of developed countries. This can discourage and frustrate researchers and simultaneously damage the research system. For instance, if being a scientist in Uruguay is rewarded with low marks in the evaluation process on the basis of universalist criteria, and possibilities to conduct research abroad appear, the loyalty to the country that makes researchers stay and build capacities in Uruguay may weaken severely. This issue was explicitly recognized by PEDECIBA when almost 20 years after its implementation it reflected on evaluation practices:

In the evaluation of researchers, it is difficult to compare merits generated in our country with those originated in countries with higher levels of scientific development. This results from the fact that producing scientific products faces more difficulties in our country. The comparison is easier if when we assess quality we weight originality, deepness and scientific rigor over the utilization of brand new sophisticated methods. The number of publications and the use of journal impact factors are distorting factors that particularly affect this comparison (PEDECIBA 2004, own translation).

A second tension is associated with fulfilling contradictory expectations. Solving local problems as it is voiced by governmental officers and enterprise organizations leaders and obtaining a good harvest of papers in academically recognized journals to raise the academic quality of national research may turn out to be mutually incompatible. The universalist type of evaluation equates high quality research with research producing articles for top ranked international journals. According to this notion, a robust community will show a research output with a good record in high impact factor journals. In order to meet this goal, it will be unlikely that a large proportion of research focuses on local topics of a limited interest for academic

mainstreams.⁵ In fact, meeting the goal may imply refraining from the main objective of developing research communities capable of addressing major contextualized problems. Such problems, by the very fact of relating to concrete and specific contexts, tend to be of a multidisciplinary nature and its thorough comprehension usually involves dialogues with actors outside academia.⁶ It is worth recalling that solving problems is always a contextual matter, for what counts as a solution in a given context may not count as such in another context.

If solving contextualized problems is part of what a country's research system fosters, a third tension may arise from the use of universalist evaluation criteria: the need to foster collective multidisciplinary efforts may clash with evaluation indicators that are mostly based on individual disciplinary performance. Institutional mandates from research organizations and universities where researchers work may promote transverse research orientations focused on particular themes that will not easily inform universalistic performance evaluation measures. Overwhelmed by the evaluation load, productivity indicators based on individual publication output have been routinely used as the best measure of individual performance. Publishing pressures have a negative impact on knowledge production genuinely oriented to interdisciplinary research and to solving problems of a complex nature (social, productive, or related to public policies) relating to particular geographical and historical conditions (Bianco et al. 2014a). Researchers may be, therefore, more willing to choose research topics well suited for the best publishing potential in the short run, or those where they can be the first author, over other research orientations that are not so promising because they require agreements with other colleagues and/or non-academic partners. Also, researchers may tend to give a second priority, in their time allocation, to those academic duties different from research which are nevertheless central for the production of knowledge but which do not directly produce tangible results. These involve participation in academic committees, administration and teaching responsibilities, and outreach among others. In order to speed up the advancement in their careers, researchers tend to develop strategies that involve selecting novel topics that can be easily published and simultaneously working at a slower pace in other themes of personal interest and/or with a higher potential for implementation of research results (Bianco et al. 2014b).

Nevertheless, on one occasion the previous tension burst out in Uruguay. It was after the first round of evaluation of the NSR, when a large proportion of researchers at INIA were not admitted into the System. The problematic issue was how researchers that were working on problems of the utmost national importance could be laid down by the System in charge of assigning academic merit. The board of the NSR reacted by organizing a special committee to deal with those researchers that did not accommodate to the universalist criteria that were utilized. We use here the

⁵ López-Piñeiro and Hicks (2015) show how decontextualized approaches in Spanish sociology are promoted by the fact that the specificities of the Spanish society are of little interest for English language audiences. Since the Spanish evaluation system emphasizes publications in high impact factor journals, these authors predict narrower and more abstract research agendas for Spanish sociology in the long run.

⁶ Chavarro et al. (2014) provide empirical evidence on the association between interdisciplinary research and research focused on local issues for the case of Colombia.

adjective “universalist” in its two meanings, relating to mainstream science and to the way some disciplines are assessed. This committee was integrated by researchers and practitioners, and its time-consuming task was to identify the merits included in non-canonical ways of communicating results and to translate them into more academic common-ground terms. INIA researchers provided most of the cases analyzed by this committee; they have an explicit mandate to solve problems which in part are posed by farmers, that is, problems of a clear contextual nature. This often leads to performing academic research, but the eagerness to transform the achieved results into academic publications is usually substantially weaker than the one an academic researcher in the same field may have. INIA authorities acknowledged the tension and were quite concerned about the way it could impact the institutional negotiation of the research agenda.

INIA researchers are shielded, at least internally, from a fourth tension that arises among some researchers in the academic milieu: the mismatch between time needed for knowledge production and research evaluation time. Several fields, especially those involving biological processes of different types, require unmovable research periods associated with biological cycles: in such cases results will be available only after these periods have elapsed. Performance research evaluation conducted on a fixed frequency, as well as the research proposal every two or three years as the main format for fund raising, enforces an artificial tyranny of time to certain research processes. Even if a typical case is that of certain agricultural science fields - in which research may take years before any result can be documented because they involve the behavior of different life forms in interaction with multiple environmental variables - there are other fields where fixed evaluation times may wrongly measure academic achievements. Examples of the latter are some branches of clinical research or environmental engineering, where laboratory work needs to be scaled-up at experimental size and then applied in real world conditions before assessing results. Uruguayan researchers, as probably most researchers in the world, resent the endless application proposals at shortly separated periods, which often implies devoting time to elaborate future proposals while the current one is still in the making or, sometimes, presenting proposals that promise results that were already obtained as a way of maximizing funding opportunities and, moreover, of getting rid of the tyranny of fixed evaluation times. A survey conducted among researchers at UdelaR in 2006 asked what the most important research policy measure for the advancement of knowledge would be; over 80% responded long-term financial support, which concomitantly implies longer periods between evaluations (Bianco et al. 2006).

The tensions just described are intertwined and with varying degrees are suffered as everyday problems by research communities everywhere. But in developing countries, and certainly in Uruguay, such tensions are expressed in a more extreme way. A small and weak research community that needs to become part of the world academy, not only as a quality audit measure but to be able to interact with other scholars at a global scale, will need to devote all its strength to that purpose. But at the same time this research community is supported by a society facing specific challenges and expecting in return of this support - however insufficient it may be - an important commitment to addressing them. They imply complex processes

of multidisciplinary and multi-level dialogues that are time consuming. Researchers know that they will be evaluated through universalist criteria, so they will try by all means to demonstrate that they belong to the world academy, which obviously is particularly difficult because they are surrounded by all types of scarcities. Therefore, chances are that the tensions are resolved against contextualized research agendas. The statement made by Hess about the importance of research evaluation for the research that is effectively done, seems after this analysis even more convincing than in the Introduction.

How to perform research evaluations became a topic of debate particularly at the stage of re-building the Uruguayan research system after 1985. PEDECIBA required criteria for the assessment of researchers at different academic levels to be included in the program. This involved agreeing upon criteria to evaluate biologists, mathematicians, chemists, computer scientists, and physicists. Evaluation processes involved researchers conducting research in different institutions (universities and research centers) and a lot of debate and concerns recognizing that all cognitive orientations differ but also highlighting that “being a good researcher” necessarily implied common traits among all fields of knowledge. Much of what was proposed a decade ago at PEDECIBA has been voiced recently by manifestos as DORA (2012) or Leiden (Hicks et al. 2015):

We suggest emphasizing quality over quantity. Evaluation based on the number of publications promotes practices that, in the long run, end up deteriorating the quality of publications. Attempting to maximize the number of publications stimulates the submission of unfinished contributions, skipping required controls, or fragmenting a contribution into smaller articles. Further, it promotes authorship strategies among researchers that appear as coauthors of all articles produced by all team members, even those publications in which some researchers have not contributed. Evaluation criteria must contribute to dissuade the researcher to make these malpractices (PEDECIBA 2004, own translation).

The Dilemma of a Negotiated Agenda

Another type of current concern in Uruguay relates to the particularities of a negotiated research agenda, that is, research that involves dialogues between scientists and lay persons at least during the identification of research problems. Non-scientists involve a wide audience exhibiting a variety of cognitive competencies and skills. That audience may include public policy representatives, social organizations, and productive sectors, among others. Some particular traits of this type of research include: (1) it takes longer to develop a working strategy because, often, the relation between user and producer of knowledge involves developing mutual trust and a common language; (2) it is local and exploratory in nature because the research problem is either new or needs to be addressed from a new perspective in order to be solved, in which case it is not part of any well-established research program; (3) it often requires much more plural approaches than those needed to solve disciplinary puzzles. Based on these characteristics, projects

developed out of a negotiated research agenda show substantial differences from the working standards in fields where research agendas are mostly set by internal influences, or where non-scientists have minimal participation. When research evaluation systems apply exactly the same universalist criteria to assess researchers who conduct research on problems from a negotiated agenda and to more disciplinary-based researchers, performance evaluation results will negatively affect the former. This aspect has been put forward by Uruguayan social scientists as well as by researchers in technological fields.

The hard core of the discussions relates to the characterization of what a researcher is and how such identity should be recognized. Understandably, the main concern of researchers is to continue being researchers. For this purpose, researchers need resources to perform research, but they also need to fulfill the demands of the reputation system (Whitley 1984) which recognizes them as researchers. Both in developed and developing contexts, the demands of a reputational system should not be unilateral if a negotiated research agenda is wanted. Room for negotiated research agendas will not be available if only a certain type of evaluation criteria prevails. Such research agendas often need longer learning processes, interactions with different stakeholders, and development of ties of mutual trust among scientists from different fields and between them and non-scientists. It is unlikely that this type of research can meet the requirements from universalist evaluation systems in terms of providing immediate publishable results.

The increasing reputational weight of the NSR fostered another debate among Uruguayan researchers, mainly at UdelaR and INIA. It revolves around the multiple tasks researchers belonging to these institutions must perform, among which research is a central responsibility but not the only one. Undergraduate teaching, meeting with farmers, participating in institutional development processes, preparing teaching materials, performing extension and outreach activities constitute a set of academic duties of fundamental importance which are usually not considered by evaluation criteria. On the basis of unilateral evaluation systems, disincentives for all these activities may end up undermining the foundations of the institutions in which most research is carried out.

We will now make a risky statement without empirical evidence, followed by an objective assertion: the NSR offers a stronger reputational system than the diverse institutional evaluation systems at place, while the latter involve economic incentives much more significantly than the former. It is therefore reasonable that researchers want to combine their institutional and their NSR belongings. However, in order to make this possible, research criteria should be harmonized because as they stand today, fulfilling both sets of demands and criteria may be unfeasible. Discussions around how such harmonization should be done have already started. INIA boosted the organization of an inter-institutional workshop on “Evaluation of researchers and research teams for innovation and development: towards a multidimensional approach” in 2013. After this workshop, a discussion group was created in which representatives of the university and several national research organizations exchanged information regarding evaluation criteria and processes and shared expectations regarding the construction of a national system of academic evaluation in which the demands of different institutions are reflected. The task

ahead is not easy; raising a voice against the universalist and quantitatively biased evaluation system of the NSR may be seen and interpreted as a claim for “soft” and permissive assessment criteria. For this reason, the current international movement of concern around what the research evaluation system is doing to the research enterprise is particularly welcome: academic legitimacy is becoming much more plural nowadays.

Final Remarks

There is no evident conclusion for a paper on such an intricate subject. We would like to think that the assertion made in the Introduction regarding the fact that academic evaluation influences the type of research that gets done has been justified in the previous sections and well exemplified for the Uruguayan case. Perhaps, it could also be considered justified the notion that the research orientation exerts an influence on the probability that research results and research capacities are put into use in development processes. The paper left an important issue without consideration, though: the more general conditions that make it possible for research to be used in development processes have not been addressed. These conditions, intimately related to the productive structure and the knowledge demand stemming from the public sphere, are outside the academic system and therefore out of the reach of researchers. After all, an improved research evaluation system better suited to promote research on development issues will be insufficient if knowledge is not actually used. In the absence of strong knowledge demands from development agents it is difficult to alter prevailing evaluation practices that mainly respond to their own logic. However difficult it may be, it is important to keep on searching for research evaluation systems able to encourage more inclusive and contextualized research agendas where all knowledge needed to foster development processes can have a space to grow in quality and in scope.

References

- Alberts, Bruce. 2013. Impact factor distortions. *Science* 340: 787.
- Arocena, Rodrigo, and Judith Sutz. 2010. Weak knowledge demand in the South, learning divides and innovation policies. *Science and Public Policy* 37(8): 571–582.
- Barré, Remi. 2010. Towards socially robust S&T indicators: indicators are debatable devices, enabling collective learning. *Research Evaluation* 19(3): 227–231.
- Benninghoff, Martin, and Dietmar Braun. 2010. Research Funding, Authority Relations, and Scientific Production in Switzerland. In *Reconfiguring Knowledge Production: Changing Authority Relationships in the Sciences and their Consequences for Intellectual Innovation*, eds. Richard Whitley, Jochen Gläser, and Lars Engwall, 81–108. New York: Oxford University Press Inc.
- Bensusán, Graciela, Natalia Gras, Daniel Inclán, Carlos E. Rodríguez, Giovanna Valenti, and Gonzalo Varela. 2014. *Reflexiones sobre la evaluación a los investigadores: una mirada desde diferentes perspectivas*. http://www.foroconsultivo.org.mx/libros_editados/evaluacion_de_la_evaluacion_subgrupos_individuos.pdf. Accessed 27 May 2015.
- Bernal, John D. 1994. *Historia Social de la Ciencia*. Barcelona: Península Ed.
- Bianco, Mariela, Natalia Gras, and Judith Sutz. 2014a. Reflexiones sobre la práctica de la evaluación académica. In *Análisis y reflexiones sobre 20 años de políticas de investigación en la Universidad de*

- la República: aciertos, dudas y aprendizajes*, eds. Mariela Bianco, and Judith Sutz, 209–235. Montevideo: CSIC-UdelaR, Trilce.
- Bianco, Mariela, Maria Goñi, and Cecilia Tomassini. 2014b. Señales transmitidas por el sistema de fomento a la investigación: tensiones en la orientación de la producción de conocimiento y las carreras académicas en Uruguay. *REDES* 20(39): 159–182.
- Bianco, Mariela, Carlos Bianchi, Andrea Bielli, Claudia Cohanoff, Ana Laura de Giorgi, Natalia Gras, and Judith Sutz (coordinator). 2006. Pensando el Plan Estratégico Nacional en Ciencia, Tecnología e Innovación. Document for the First PENCTI Workshop, CSIC-UdelaR. http://csic.edu.uy/renderPage/index/pageld/275#heading_892. Accessed 18 March 2016.
- Bunders, Joske. 1987. The practical management of scientists' actions: the influence of patterns of knowledge development in biology on cooperations between university biologists and non-scientists. In *The Social Direction of the Public Sciences*, eds. Stuart S. Blume, Joske Bunders, Loet Leydesdorff, and Richard Whitley, 39–74. Dordrecht: Reidel Publishing Co.
- Bunders, Joske. 1990. *Biotechnology for small-scale farmers in developing countries. Analysis and assessment procedures*. Amsterdam: VU University Press.
- Bunders, Joske, and Jacqueline Broerse. 1991. *Appropriate biotechnology in small-scale agriculture: How to orient research and development*. Wallingford: CAB International.
- Chavarro, Diego, Puay Tang, and Ismael Rafols. 2014. Interdisciplinarity and research on local issues: Evidence from a developing country. *Research Evaluation* 23(3): 195–209.
- Dahler-Larsen, Peter. 2014. Constitutive Effects of Performance Indicators: Getting beyond unintended consequences. *Public Management Review* 16(7): 969–986.
- de Jong, Stefan P.L., Pleun van Arensbergen, Floortje Daemen, Barend van der Meulen, and Peter van den Besselaar. 2011. Evaluation of research in context: an approach and two cases. *Research Evaluation* 20(1): 61–72.
- De la Mothe, John, and Gilles Paquet. 1996. *Evolutionary Economics and the New International Political Economy*. London: Pinter.
- Dobbins, Michael, and Christoph Knill. 2014. *Higher Education Governance and Policy Change in Western Europe. International Challenges to Historical Institutions*. London: Palgrave Macmillan.
- DORA. 2012. San Francisco Declaration on Research Assessment. <http://am.ascb.org/dora/>. Accessed 20 June 2015.
- Dicyt-MEC. 2012. *Informe a la Sociedad. Ciencia, Tecnología e Innovación en Uruguay en los últimos años*. Montevideo: Ministerio de Educación y Cultura.
- Elzinga, Aant. 1988. The consequences of evaluation for academic research. *Science Studies* 1: 5–14.
- Evidence Ltd. 2007. The use of bibliometrics to measure research quality in the UK higher education institutions. <http://www.universitiesuk.ac.uk/highereducation/Documents/2007/Bibliometrics.pdf>. Accessed 20 August 2015.
- Gläser, Jochen, and Grit Laudel. 2016. Governing Science. *European Journal of Sociology* 57(01): 117–168.
- Gläser, Jochen, and Grit Laudel. 2007a. The social construction of bibliometric evaluation. In *The Changing Governance of the Sciences. The Advent of Research Evaluation Systems*, eds. Richard Whitley, and Jochen Gläser. Dordrecht: Springer.
- Gläser, Jochen, and Grit Laudel. 2007b. Evaluation Without Evaluators: The Impact of Funding Formulae on Australian University Research. In *The Changing Governance of the Sciences. The Advent of Research Evaluation Systems*, eds. Richard Whitley, and Jochen Gläser, 127–151. Dordrecht: Springer.
- Hemlin, Sven, and Soren Barlebo Rasmussen. 2006. The Shift in Academic Quality Control. *Science, Technology, & Human Values* 31(2): 173–198.
- Hess, David. 2007. *Alternative Pathways in Science and Industry. Activism, Innovation, and the Environment in an Era of Globalization*. Cambridge: The MIT Press.
- Hessen, Boris. 1931. "The Social and Economic Roots of Newton's Principia", paper presented at 2nd International Congress of the History of Science and Technology. London June 20–July 3.
- Hicks, Diana. 2004. The Four Literatures of Social Science. In *Handbook of Quantitative Science and Technology Research*, eds. Henk Moed, Wolfgang Glänzel, and Ulrich Schmoch, 473–496. Dordrecht: Kluwer Academic Publishers.
- Hicks, Diana. 2006. The Dangers of Partial Bibliometric Evaluation in the Social Sciences. *Economia Politica XXIII* 2: 145–162.

- Hicks, Diana. 2013. One size doesn't fit all: On the co-evolution of national evaluation systems and social science publishing. *Confero: Essays on Education Philosophy and Politics*. doi:10.3384/confero13v1121207b
- Hicks, Diana, Paul Wouters, Ludo Waltman, Sarah de Rijcke, and Ismael Rafols. 2015. The Leiden Manifesto for research metrics. *Nature* 520(April): 429–431.
- Katz, J. Sylvan. 1999. Bibliometric Indicators and the Social Sciences. ESRC <http://users.sussex.ac.uk/~sylvank/pubs/ESRC.pdf>. Accessed 20 August 2015.
- Kuhn, Thomas. 1970. *The Structure of Scientific Revolutions*, 2nd ed. Chicago: University of Chicago Press.
- López-Piñero, Carla, and Diana Hicks. 2015. Reception of Spanish sociology by domestic and foreign audiences differs and has consequences for evaluation. *Research Evaluation* 24(1): 78–89.
- Martin, Ben, and Richard Whitley. 2010. The UK Research Assessment Exercise: A Case of Regulatory Capture? In *Reconfiguring Knowledge Production: Changing Authority Relationships in the Sciences and their Consequences for Intellectual Innovation*, eds. Richard Whitley, Jochen Gläser, and Lars Engwall, 51–79. New York: Oxford University Press Inc.
- McNay, Ian. 2009. Research Quality Assessment: objectives, approaches, responses and consequences. In *Academic Research and Researchers*, eds. Angela Brew, and Lisa Lucas, 35–53. London: McGraw Hill.
- Meier, Frank, and Uwe Schimank. 2010. Mission Now Possible: Profile Building and Leadership in German Universities. In *Reconfiguring Knowledge Production: Changing Authority Relationships in the Sciences and their Consequences for Intellectual Innovation*, eds. Richard Whitley, Jochen Gläser, and Lars Engwall, 211–237. New York: Oxford University Press Inc.
- Merton, Robert. 1988. The Matthew effect in science. *ISIS* 79(4): 606–623.
- Merton, Robert. 1942. The Normative Structure of Science. In *The Sociology of Science*. Chicago: University of Chicago Press.
- Morris, Norma. 2010. Authority Relations as Condition for, and Outcome of, Shifts in Governance: The Limited Impact of the UK Research Assessment Exercise on the Biosciences. In *Reconfiguring Knowledge Production: Changing Authority Relationships in the Sciences and their Consequences for Intellectual Innovation*, eds. Richard Whitley, Jochen Gläser, and Lars Engwall, 239–264. New York: Oxford University Press Inc.
- PEDECIBA (Programa de Desarrollo de las Ciencias Básicas). 2004. Criterios, herramientas y procedimientos generales para la evaluación de la actividad académica de los investigadores. <http://www.pedeciba.edu.uy/docspd/CritEvalInv04.pdf>. Accessed 21 Aug 2015.
- Power, Michael. 1987. *The Audit Society. Rituals of Verification*. Oxford: Oxford University Press.
- Regeer, Barbara, Anne-Charlotte Hoes, Mariette van Amstel, Francisca Caron-Flinterman Saane, and Joske Bunders. 2009. Six Guiding Principles for Evaluating Mode-2 Strategies for Sustainable Development. *American Journal of Evaluation* 30: 515–537.
- Rosenberg, Nathan. 1982. *Inside the Black Box*. Technology in Economics. Cambridge: Cambridge University Press.
- Sábato, Jorge, and Natalio Botana. 1968. La ciencia y la tecnología en el desarrollo futuro de América Latina. *Revista de la Integración* 1(3): 15–36.
- Sahel, Jose Alain. 2011. Quality versus quantity: assessing individual research performance. *Science Translational Medicine*. doi:10.1126/scitranslmed.3002249.
- Schimank, Uwe. 2005. “New Public Management” and the Academic Profession: Reflections on the German Situation. *Minerva* 43: 361–376.
- UA-CSIC (Unidad Académica de CSIC). 2003. *Grupos de Investigación en la Universidad de la República*. Montevideo: CSIC-UdelaR.
- van Dalen, Hendrik, and Kéne Henkens. 2012. Intended and Unintended Consequences of a Publish-or-Perish Culture: A Worldwide Survey. *Journal of the American Society for Information Science and Technology* 63(7): 1282–1293.
- van der Most, Frank. 2010. Use and non-use of research evaluation: A literature review. Centre for Innovation, Research and Competence in the Learning Economy (CIRCLE), Lund University, Working Paper. http://wp.circle.lu.se/upload/CIRCLE/workingpapers/201016_vanderMost.pdf. Accessed 20 August 2015.
- Van Raan, Anthony F.J. 1996. Advanced bibliometric methods as quantitative core of peer review based evaluation and foresight exercises. *Scientometrics* 36(3): 397–420.

- Vessuri, Hebe, Jean-Claude Guédon, and Ana María Cetto. 2014. Excellence or quality? Impact of the current competition regime on science and scientific publishing in Latin America and its implications for development. *Current Sociology* 62(5): 647–665.
- Whitley, Richard. 1984. *The Intellectual and Social Organization of the Sciences*. Oxford: Oxford University Press.
- Whitley, Richard. 2007. Changing Governance of the Public Sciences: The Consequences of Establishing Research Evaluation Systems for Knowledge Production in Different Countries and Scientific Fields. In *The Changing Governance of the Sciences. The Advent of Research Evaluation Systems*, eds. Richard Whitley, and Jochen Gläser, 3–27. Dordrecht: Springer.
- Whitley, Richard. 2010. Reconfiguring the Public Sciences: The Impact of Governance Changes on Authority and Innovation in Public Science Systems. In *Reconfiguring Knowledge Production: Changing Authority Relationships in the Sciences and their Consequences for Intellectual Innovation*, eds. Richard Whitley, Jochen Gläser, and Lars Engwall, 3–47. New York: Oxford University Press Inc.
- Ziman, John. 1994. *Prometheus Bound. Science in a Dynamic Steady State*. Cambridge: Cambridge University Press.