

National Research Evaluation Systems and the Social Sciences

Michael Ochsner¹ Ginevra Peruginelli²

¹ochsner@evalhum.eu

ETH Zurich, Zurich (Switzerland)

FORS, Lausanne (Switzerland)

<https://orcid.org/0000-0002-1758-9590>

²peruginelli@igsg.cnr.it

Istituto di Informatica Giuridica e Sistemi Giudiziari (IGSG - CNR), Florence (Italy)

<https://orcid.org/0000-0002-9331-4476>

Chapter accepted for publication in Engels, T. C. E. & Kulczycki, E. (Eds.). (2021). *Handbook on Research Assessment in the Social Sciences*. Edward Elgar, ISBN 9781800372542.

Abstract

Many countries have introduced national evaluation exercises. The methods commonly applied have been developed for the natural and life sciences and are criticized for being ill-adapted to the social sciences. After giving an overview of different classifications of countries' evaluation procedures, we present an empirically derived typology of national research evaluation systems that takes into account whether adaptations to disciplinary specificities, such as social sciences knowledge production and dissemination practices, are considered. The analysis shows that national evaluation systems are diverse and complex combinations of research evaluation practices at the national, regional and institutional level, including the influence of the international level on research funding of projects and careers. We conclude that diversity rather than standardisation in evaluation practices is key to reach the aim of evaluation: to provide researchers and university administrators with a tool to improve their organisation and research practice.

Keywords: research evaluation, evaluation systems, country comparison, social sciences and humanities, public management, performance-based research funding

Introduction

Evaluation is innate in research. In fact, it is a constituent part of the research process: without evaluating research, new research cannot be conducted. At the same time, evaluation can be separated from the research process, as is the case in systematic research evaluation. Consequently, evaluation can be performed for several goals, on various levels and in different time frames by various actors, leading to complex systems of evaluation procedures. Evaluations in general are social processes and cannot be separated from the social, political and historical contexts because evaluations serve as a form of negotiation of values, a process of sense making and self-reflection of an organisation in society (Dahler-Larsen, 2012, pp. 16-17, 234). We argue that this social construction of research evaluation procedures and their combinations is neglected too often in the current international debate

on evaluation criteria, indicators and methods, which focuses rather on general aspects of science and looks for “the” optimal indicator or evaluation procedure or on general principles like open science (exceptions are, for example, Barré, 2010; Bianco et al., 2016; De Rijcke et al., 2019). Research evaluation, however, takes place in a political-cultural context and should, in our opinion, reflect disciplinary practices; it should represent the link between the science system and science policy.

In the last decades, research evaluation procedures have been affected by significant changes, due to new policies of public management and increasing pressure on efficiency and accountability as well as technological advances. These developments have been widely considered as evidence of the intrusion of managerial governance into university procedures and organisation, shifting from process to output control (Deem et al., 2008; Torrance, 2020). Researchers perceive evaluation as a management tool disconnected from research practice while they do not relate the evaluative nature of the research process itself to the term “evaluation”, taking a critical stance towards systematic evaluation. At the same time, governments around the world have introduced various forms of research evaluation procedures based on ideas of management in business, increasing the frictions between university management and researchers (Rolfe, 2013).

The first disciplines to adopt systematic procedures and methods for the assessment of research have been subfields of the natural and life sciences, such as biochemistry, applied physics and electronic and mechanical engineering. The SSH are, also, increasingly subject to these evaluation practices (Burrows, 2012; Guillory, 2005). Of course, the evaluation procedures are adapted to and developed for these fields from the natural and life sciences and are heavily based on bibliometrics, which have been proven inadequate for the SSH disciplines (see, e.g., Hicks, 2004; Larivière et al., 2006; Nederhof, 2006), but also for many fields that fall under the STEM umbrella, such as mathematics or engineering. As a consequence, a relevant part of SSH scholars shows a hostile attitude to assessment practices even questioning the value of research evaluation procedures as such or refusing them altogether (see, i.e., Lock & Martins, 2011; Plumpe, 2009; Sievers, 2008).

Understanding how SSH research is evaluated across countries is crucial to figure out this sort of opposition. Such an investigation could contribute to design evaluation procedures that enhance SSH research and hopefully reach acceptance among SSH scholars.

While a number of overviews and typologies are available on national evaluation procedures (Coryn et al., 2007; Hicks, 2010; 2012; Jonkers & Zacharewicz, 2016; Lepori et al., 2007; Lepori et al., 2017; Geuna & Martin, 2003; Von Tunzelmann & Kraemer Mbula, 2003) and representative associations of universities and other R&I actors have worked hard to improve the knowledge base for the discussion (Saenen et al., 2019; Science Europe, 2020), there is still poor knowledge of how research is evaluated in different countries, especially with regard to the SSH. Most studies only cover a restricted number of countries, usually those for which information is publicly available in English and where funding is attributed according to the evaluation. Furthermore, disciplinary differences are often left out of these works and research studies.

In this chapter, we propose a conceptualization of *national research evaluation systems* tackling these limitations by first proposing a typology of national evaluation systems and then presenting an inventory of national research evaluation procedures showing the diversity of evaluation procedures across Europe. Thus, we are able to balance the need to systematise and simplify complex systems to common denominators, and the necessity to understand the complexity of the systems and the reasons for, and benefits of, diversity. We develop further the work conducted within the COST-Action CA15137 “European Network for Research Evaluation in the Social Sciences and Humanities (ENRESSH)” (see also Ochsner et al., 2018; Ochsner et al., 2021) paying particular attention to the social sciences. Given that the boundaries between social sciences and humanities are fluid and definitions what disciplines belong to which class differ strongly among countries, studies on research evaluation often treat them together under the SSH umbrella. Challenges in evaluation are similar

among social sciences and humanities; therefore, the paper refers to research directed to social sciences and/or humanities and mention issues specific to social sciences when needed.

The chapter starts with a description of the several typologies proposed to systematize research evaluation procedures across countries. The discussion will then focus on the concept of diversity of research evaluation procedures. We will reflect on the relationship of national research evaluation and the social sciences as a delicate one and we will conclude arguing that evaluation systems should reflect the specific research situation of a country and the characteristics of the disciplines and that, consequently, policy makers and designers of evaluation systems should conceive diversity as an enrichment rather than a limitation.

Typologies of National Evaluation Procedures

Research evaluation is an object of study for quite some time. Three general differentiations of evaluation procedures have been put forward by scholars analysing research evaluation procedures over the years (see also Whitley, 2007): a) the stage of evaluation, b) its link to funding and c) the primary method of evaluation. The *stage of evaluation* reflects whether evaluation takes place before the research has been conducted (*ex-ante* evaluation), for example research proposals or appointment procedures, or after the research has been conducted (*ex-post* evaluation), for example institutional evaluation. The *link to funding* contrasts evaluations that allocate funding on the basis of the evaluation results (*summative* evaluation), for example performance-based research funding, against evaluation that aims at improving the research process without any financial consequences (*formative* evaluation). Finally, the primary method of evaluation differentiates regarding methods that are used to evaluate, the most evident being whether the evaluation is peer review-based or metric-based, but also the level on which is evaluated (e.g. institutions, departments or units) or who is responsible for it (e.g. the state, university or an agency) are methodological aspects of evaluation.

Systematic comparisons of how research is evaluated in different countries and typologies of national evaluation practices, however, appeared much later (Coryn et al., 2007; Galleron et al., 2017; Geuna et al., 1999; Geuna & Martin, 2003; Hicks, 2010; 2012; Jonkers & Zacharewicz, 2016; Lepori et al., 2007; Lepori et al., 2018; Ochsner et al., 2018; Von Tunzelmann & Kraemer Mbula, 2003). Such classifications differ regarding which of the three differentiations they consider in their endeavour. Most of the classifications take only one or two into account. In the following, we present a grouping of the different classifications regarding the three differentiations.

The first group of classifications has in common that they restrict the procedures they analyse to those allocating funding to institutions. They distinguish these procedures according to different methods of evaluation. Geuna and colleagues (Geuna et al., 1999; Geuna & Martin, 2001, 2003) considered only *ex-post* evaluation in their early attempt to classify different countries' approaches to research evaluation. Focussing on how funding is linked to evaluation, four groups were identified (Geuna & Martin, 1999): performance-based; indicator-based (but not research performance; often it is educational size); funding open to negotiation; research assessment and funding are separated. However, it is notable that as the study develops over time from the initial report in 1999, the classification becomes less and less defined and the number of countries included diminishes in each iteration, showing that the evaluation landscape is diverse and changes over time and that it is a continuum from *ex-post* performance-based funding allocation on one end to funding based solely on educational size on the other (Geuna & Martin, 2001, pp. 26-27, 41; Von Tunzelmann & Kraemer Mbula, 2003).

Zacharewicz et al. (2018) combine two dimensions of how funding is attributed to universities: performance-based or bulk funding and method of evaluation. From policy document analysis and expert information, they derive three types of university funding allocation: no performance-based component in funding allocation; a limited performance-based funding allocation; performance-based allocation. The latter is divided in two subtypes: metric-based and peer review-based funding allocation. They use a set of indicators grouped into five categories (education metrics, historical

continuity, bibliometrics, other (patents, knowledge transfer), peer review and performance contracts) to assign countries to the four groups based on the PREF study (Reale, 2017).

Coryn et al., (2007) also classified countries according to funding and method, however in a more sophisticated way. They differentiate three types of funding links: Type I: large-scale performance-based exercises, Type II: bulk funding, Type III: indicator-driven performance-based models. Evaluation approaches are differentiated in two groups: A, systematic and consistent; and B, pluralized approaches. The combination of those two dimensions would lead to six types of evaluation procedures used for allocating funding, however, only Type III varies in method according to the classification, thus the classification results in four types: Type I-A, large-scale exercise and consistent method; Type II-B, bulk funding pluralised approach; Type III-A, indicator-driven consistent method; and Type III-B, indicator-driven pluralised approach.

Hicks (2010; 2012), on the other hand, focuses exclusively on so-called performance-based funding models (PRFSs, i.e. models that are based on research output, thus excluding bulk-funding models or models based on input measures such as number of students) and investigates different methods to link evaluation to funding. She identifies five types of PRFSs: University-level using paper counts; university-level using paper counts and citations; department or field-level using peer review; department or field-level using informed peer review; group-level using peer review. Her analysis shows that in some countries, different procedures are in place that belong to different types or the type changes over time. Hicks concludes that PRFS designers have to face the challenge of balancing between peer review and metrics and the level on which to evaluate. Peer review is accepted among scholars while metrics seem to deliver more current results of evaluations as less time elapses between research and the evaluation. Additionally, one size does not fit all and PRFSs must accommodate different research practices across fields particularly regarding social sciences and humanities, for which the use of metrics has been proven inadequate or at least complicated.

The studies of this first group of classification suggest that in all countries having implemented a PRFS, they did so to improve research excellence. But both Hicks (2012) and Geuna and Martin (2003) question the usefulness of PRFSs to achieve this goal. They argue that PRFSs with their focus on scientific outputs are on the one hand not encouraging transfer activities (Hicks, 2012, p. 259) and on the other they are costly, and returns are diminishing after the initial increase at the time of implementation (Geuna & Martin, 2003, p. 303). The studies also show that evaluation procedures change considerably over time, while Coryn et al. (2007) found that exercises with constant methods achieved consistently the highest acceptability among experts.

The second group of classifications takes only project funding (ex-ante evaluation) into account. Lepori et al. (2007) demonstrate that the importance of competitively funded research projects increases rapidly. Studying six countries' national research policies regarding competitive project funding, they present two major findings: First, funding instruments, agencies and beneficiaries differ substantially across countries. Second, there are also commonalities: project funding is the second pillar of public funding of research and the share of competitive funding grows across all countries. Moreover, funding instruments shift to being oriented towards specific topics. Finally, the diversity in funding instruments increased in all countries but the Netherlands. Zacharewicz et al. (2018) continue the analysis including more European countries and find that in almost all countries, project funding is increasing while a considerable variation persists in the share of project funding on total research funding. However, some countries also show a decrease in the share of project funding. Unfortunately, they do not combine their analyses of PRFSs and project funding but rather combine the two by introducing an indicator for performance orientation of public R&D funding (Lepori et al., 2018). The studies in this group do not classify countries analytically but remain descriptive.

The third group takes research evaluation *systems* into the focus. It also includes formative evaluation procedures that do not (directly) link evaluation outcomes to funding. Whitley defines research evaluation systems as "organised sets of procedures for assessing the merits of research" (Whitley

2007, p. 6). He differentiates between strong and weak research evaluation systems, which fall mostly together with summative and formative evaluation procedures even though the amount of competitive project funding (i.e. subject to ex-ante evaluation) is included in his analysis. However, we argue that national research evaluation systems are not always ‘organised sets’ of evaluation procedures. If the goal of research evaluation is to improve or at least influence research practice – and, indeed, if it wasn’t the goal, why would one bother to evaluate in the first place (see Hicks, 2012; Jonkers & Zacharewicz, 2016; Zacharewicz et al., 2018) – a perspective on isolated research evaluation procedures or funding schemes from the perspective of policy makers or university management is not sufficient. Researchers work in complex and global contexts and are influenced by many evaluation procedures. Consequently, if policy aims at influencing research practice (also if only to increase excellence), the diversity of such evaluation procedures needs to be taken into account. The third group of classifications goes more into detail (Galleron et al., 2017; Van Gestel & Lienhard, 2019b; Giménez-Toledo et al., 2019; Ochsner et al., 2018). This type of classification takes the perspective from the shop floor, i.e., the researchers. What are the evaluation procedures a researcher is encountering at different levels and how does this potentially influence the researchers’ behaviour and practice? We therefore define the term *national research evaluation systems* more broadly, understanding them as actual combinations of research evaluation practices at the national, regional and institutional level, including the influence of the international level on research funding of projects and careers.

A multi-stage project comparing such complex research evaluation systems across European countries showed that there is much more complexity to the endeavour than the current studies presented above suggest: A country can have several different procedures in place and formal definitions of a procedure can differ quite substantially from the actual implementation and (local) use of the evaluation outcome. It is thus not so much a surprise that even experts disagree about how research is evaluated in their country (Galleron et al., 2017; Ochsner et al., 2018). The studies demonstrate that countries do not have a dominant procedure nor a coherent set of procedures in place. Rather, in each country, there is a complex combination of evaluation procedures, each with its own aims, scope, object and governing body (see for law studies Van Gestel & Lienhard, 2019b). While the typologies above stayed descriptive, assigning countries to theoretically pre-defined types, we chose an analytical approach. We fielded a survey among specialists in research evaluation of 32 countries, who were asked to answer questions about research evaluation procedures in their countries. We agreed upon eight main characteristics of national evaluation systems and used multiple correspondence analysis to derive a two-dimensional space of national evaluation systems. The first dimension considered metric aspects of evaluation whereas the second dimension reflected whether disciplinary differences, i.e., adaptations to SSH fields, are respected. Using the two dimensions to structure how countries’ evaluation systems relate to the eight characteristics, we empirically derived five ideal types of national research evaluation systems (Ochsner et al., 2018): 1) “no national database, non SSH-specific” (not having a national publication database, using mainly non-metric evaluation procedures and not allowing for SSH adaptations); 2) “non-metric, SSH-specific” (not having a publication database, not relying on metrics for their evaluations, not incentivising publications in English and having dedicated funding programs for SSH disciplines); 3) “performance-based funding, mixed methods” (having a PRFS in place that allows for SSH-specific adaptations and is based on metrics derived from a national publication database, but adding peer review practices to complement the metric system); 4) “performance-based funding, metric” (PRFSs being based on a national database and a metric evaluation that allows for SSH adaptations and not incentivising publications in English); 5) “metric, push for English” (metric evaluation based on a national publication database linked to funding and not allowing for SSH adaptations and incentivising publications in English). These five ideal types, schematically depicted in Figure 1, are abstract representations of evaluation systems derived from actual evaluation systems. Real national evaluation systems are rather combinations of several of the five ideal types. Note that this approach is the opposite of the other typologies: while the other typologies classify countries into theoretically

pre-defined clusters aiming to describe the evaluation systems, this approach takes evaluation systems and analyses their relation regarding eight characteristics to derive ideal types describing how evaluation systems can look like. This analytical approach helps us understand what the relevant differences in complex systems are rather than trying to put national systems into a procrustean bed of predefined boxes. The goal of the typology is thus not to put a country in one type but rather to understand conceptual differences in evaluation systems. In other words, we are aiming at analysing what kind of evaluation systems exist and in what characteristics they differ. Thus, instead of looking at a country, putting it into a cluster and argue that the cluster represents how research is evaluated in a single country, we aim at drawing a map of ways how research evaluation is organised in countries. For example, we find a remarkable regional clustering of countries according to these types: Southern European countries are closer to type 1 and Eastern European countries affiliating with types 3 and 4 (Ochsner et al., 2018). The typology thus helps us to see that Eastern European countries are concerned with different issues than Western European countries. They are searching to integrate their research more into the existing international landscape of research and thus put emphasis on English language publications and do not focus on SSH research to achieve this goal. However, how this is implemented in each country might differ. Further, we learn that there are more aspects relevant to characterising national evaluation systems as the known aspects of summative vs. formative, metric vs. peer review or ex-ante vs. ex-post. Interestingly, we notice that on the metric dimension, the fact of having a national research information system is more relevant to distinguish between evaluation systems than having implemented a performance-based funding system. Also, national evaluation systems differ in how important English as publication language is and whether discipline-specific evaluation methods are implemented.

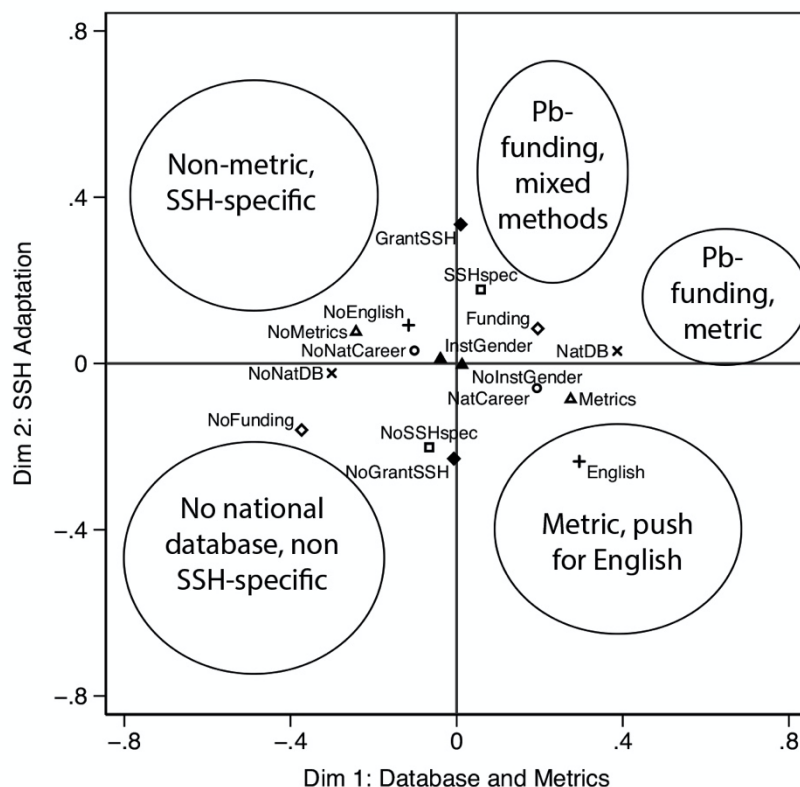


Figure 1. Schematic presentation of the five ideal types identified in Ochsner et al. (2018).

Note that for an easier interpretation, the orientation of the axes were switched compared to the original publication. English/NoEnglish: system incentivises (or not) English language publications; (No)Funding: evaluation results affect funding; (No)GrantSSH: SSH-specific grant programmes; (No)InstGender: evaluation procedures reflect gender issues; (No)Metrics: main method of evaluation are metrics; (No)NatCareer: national career promotion procedure; (No)NatDB: national publication database; (No)SSHspec: SSH-specific institutional evaluation procedures.

The Diversity of National Research Evaluation Systems

While typologies help to understand commonalities among different countries and to find regularities between contexts and their impact on evaluation, they necessarily simplify the complex nature of the research evaluation systems consisting of many actors, such as governments, funders, learned societies, universities, researchers. The typology by Ochsner et al. (2018) allows a first glance at the diversity of research evaluation practices in several European countries, as countries spread widely in a two-dimensional space and countries within one ideal type differ on many aspects. In the follow-up project, experts from the countries collected information on their countries' regulations for evaluation procedures, which showed that the three evaluation procedures used in the project did not suffice for a detailed representation of the evaluation systems. In an iterative process, seven types of procedures were identified: accreditation, institutional evaluation (divided in performance-based and formative evaluation), career promotion, excellence initiative, project funding and evaluation of academies of sciences. Each of these procedures, or their inexistence, impacts how research in a country is conducted or organised. In the following, we describe these types of evaluation procedures and then present how countries differ in their compositions of types of evaluation procedures that make up the specific research evaluation system of the country.

Types of evaluation procedures

Accreditation is defined as the evaluation procedure an institution has to go through to be acknowledged as a university. Usually, this means that research and teaching is evaluated on the institutional level in order to allow the institution to call itself a university. Often, researchers are not really affected by this evaluation procedure and do not even know about it as it concerns the administration of the university as a whole and thus it's mainly the persons involved in university management, such as directors of institutes or deans who are concerned with it.

Institutional evaluation concerns how research of departments, institutes or chairs are evaluated within an institution. Sometimes, the evaluation at this level follows the same procedure throughout the country thus also evaluating research across institutions, sometimes each institution has its own way of evaluating research, sometimes there are combinations of the two. This is one of the most important evaluation procedures because it is all about how researchers are evaluated at their workplace. As there is quite some diversity in how this evaluation is organised, we distinguish between two forms of institutional evaluation: performance-based and formative evaluation. Some countries have only one of the two forms implemented, while others apply the two forms at the same time. Many countries do not have national procedures for institutional evaluation, but the research evaluation of chairs or individuals is rather organised at the regional or even institutional level.

The national career promotion offers nationally standardised promotion opportunities to academics taking into account individual career profiles and paths. In some countries, researchers have to go through a national procedure to be promoted to the next career level, in most cases this applies to the so-called habilitation procedure, i.e., the degree that enables a researcher to become a professor. Sometimes, researchers even have to go through a centralised national procedure for each step in their career, e.g., from assistant to associate professor and from associate to full professor. Characteristic for this procedure is the centralised form it takes: a national institution is responsible for attributing the degree to researchers. Of course, if such a national procedure for career promotion is in place, this has quite some importance for the researchers. They usually know what is expected from them. If it is a one-time decision, it will lose its effect as soon as the degree is awarded, if it is a continuous process, it has a lasting effect on researchers. Habilitation procedures that are organised at the institutional level or even at the level of department or faculty are not considered in this group.

The "excellence initiative" procedure serves the goal of identifying excellent institutions in a country. It distinguishes high-level research-centred institutions from the more ordinary universities that fulfil other missions than excellence. This policy measure is introduced to mimic the ivory league

universities in the USA and often to counterbalance the upgrading to university level of institutions that focus on applied sciences.

National project funding includes procedures for competitive project financing organised by a national research funding institution. In many countries, there is also substantial project funding coming from private institutions, often in the form of foundations; here, we only take national project funding institutions into account that are distributing governmental money.

Evaluation procedures for academies of science entail procedures in which research conducted within national academies of sciences are evaluated. Sometimes, historically, universities were institutions of teaching and not of research. Research has taken place at academies of sciences. If research at such academies of sciences is evaluated in a separate, nationally organised evaluation, this procedure applies. Often, academies of sciences are rather umbrella organisations of learned societies or institutions engaged in science policy rather than in research. In such cases, this procedure does not apply. Also, if researchers at academies of sciences are evaluated following the same procedure as researchers at universities, the evaluation procedure falls under the corresponding general institutional evaluation procedure.

The Countries' Composition of Research Evaluation System

While a more interpretative typology of evaluation procedures is still work in progress, Table 1 already provides valuable insights about the diversity of national evaluation systems across Europe. Columns represent the seven types of evaluation procedures, while rows identify countries. Grey-shaded cells indicate that this type of procedure is implemented in the country. As we do not have a specific theory how types of procedures are implemented, we ordered the countries alphabetically. Table 1 shows that in all countries, governments assign some of their funding to projects distributed competitively, sometimes according to thematic calls but also including blue-sky research. Almost all countries do have an accreditation procedure in place that defines whether an institution can call itself "university". Five countries do only have accreditation and project funding in place (AT, BE-WBF, CH, CY, IE). On the other side of the scale, three countries have six out of the seven procedures in place (BG, ES, UK), but in different combinations: Bulgaria does not have a formative national evaluation in place, Spain does not have academies of science and the UK does not have a centralised institution for career promotion. In-between the two groups, we have all kinds of combinations of evaluation procedures. We also see that national institutional evaluation is more often performance-based (16 countries out of 24) than formative (11 countries), however, eight countries combine a performance-based procedure with a formative procedure, a fact that is rarely taken into account in the scholarly discussion on research evaluation. This might be due to the fact that most typologies focus on performance-based procedures only. We also learn from the Table that not many countries have implemented an evaluation procedure for academies of sciences (or other research institutions outside of university); this has two reasons, first, not all countries have academies of sciences that conduct research and second, sometimes the academies of sciences, when conducting research, are evaluated according to the same procedure as universities. Finally, only a few countries have a centralised institution for career promotion. While only Southern and Eastern European countries have such an evaluation procedure in place, not all Southern and Eastern European countries do.

Of course, all national systems outlined here are situated in the European context and there is no doubt that research evaluation practices on the European level influence the strategic behaviour of researchers, and therefore also affect national evaluation frameworks (Snel, 2019). Several are the EU funding instruments¹ playing an important role in influencing national evaluation procedures. Some of them focus on researchers' excellence (ERC Grants, Marie Curie Actions, Max Weber

¹ European research policy has a basis in Title XIX of the Treaty on the Functioning of the European Union. See Art. 179(1) and 182(1)

Fellowships) while others are dedicated to support partnerships between EU countries, the private sector and other stakeholders (Horizon Europe).

Table 1. Types of evaluation procedures and countries' composition of research evaluation system

Country	Accreditation	National Evaluation: formative	National Evaluation: performance-based	National Excellence Initiative	National Career Promotion	Project Funding by Government	Evaluation for Academies of Sciences
AT							
BE-WBF							
BG							
CH							
CY							
CZ							
DE							
ES							
FI							
FR							
HR							
IE							
IL							
IT							
LT							
LV							
NO							
PL							
PT							
RO							
SI							
SE							
UK							
ZA							

All these schemes are addressed to all European researchers and in some cases to non-European countries' researchers. This is also in line with the nature of scientific research which is supranational. Scholarship is generally international and, at least at an earlier career stage, scholars might look across borders. As a consequence, national evaluation systems are not isolated systems (Snel, 2019) but embedded in a supranational research community and the European level should not be forgotten.

The Commission Recommendation of 11 March 2005 on the European Charter for Researchers (European Commission, 2005) represents a generic invitation to European Member States to establish a level of uniformity on research assessment procedures across countries. While standardisation of research assessment has not taken place for various reasons, relevant progress has been made in removing the geographical barriers to researchers' mobility and the fragmentation of research careers in Europe, driven by the European Charter and the Code of Conduct for the Recruitment of Researchers already adopted by more than 1200 organizations. A direct outcome of the Charter is represented by the adoption of the "HR Excellence in Research" label, which introduces the concepts of open, transparent and merit-based recruitment at many institutions. This trend to fostering researcher mobility as well as skill and career development opportunities within the EU is confirmed by more recent policy documents such as the Communication on "A new Era for Research and Innovation" of 30 September 2020 (European Commission, 2020).

In this context it is worth to mention the European Framework for Research Careers (European Commission, 2011), which represents a voluntary transparency instrument intended to make research career structures generally comparable across countries. According to the European Framework, researchers across Europe fall into defined categories. However, when put into national contexts, these categories remain not easily comparable research career structures and despite there being cross-country and cross-sector mobility, many barriers remain. In such a context, scholars have to cope with this diversity and need to be aware of different evaluation approaches (career systems, career steps and progression). If they want to move from one country to another, they should know that they need to follow a strategy that is compatible with many systems. This implies that a global model in academic career promotion and a development of a harmonized academic identity should be desirable. This does not mean standardization, or even worse, a unification of research evaluation systems, but a harmonization of fundamental values which should be coherent across Europe (for example: fields of science, career stages).

Whereas the European Union introduces a supranational level of evaluation practices, scientific careers are international going beyond the EU. However, national higher education systems differ structurally and even increasing importance of internationality did not erase such “historically derived arrangements” (Clark, 1978, p. 242; see also Whitley, 2003). Consequently, and as experienced in the discussions in the work group on national evaluation systems, scholars are still only weakly aware of differences in national evaluation systems. Even international scholars specialised in evaluation are often framed in the national contexts and discourse of research evaluation. For example, Dutch and English scholars will more likely be convinced that the only really relevant criterion in the end is societal impact while Swiss and Irish scholars might put forward different aspects of research and the danger of being put in a procrustean bed of criteria not related to the relevant subjects of research because research aims at solving or prevent problems in the future not those from the past; focussing on direct social impact will, according to this view, only lead to science answering questions from the past. Similarly, some scholars will emphasise the relevance of publications in international databases and citations derived from these, while again others will argue that without national databases, no evaluation can seriously be accomplished. Despite international scholarship, researchers are influenced by the national evaluation procedures in how they evaluate others (see also Aagaard, 2015, Reymert et al., 2020). Therefore, it is not a big surprise that an international group of evaluation researchers needed several years to agree on what national research evaluation means. Table 1 reminds us of the diversity of what scholars speak about when they talk about evaluation. It is rarely the case that the term evaluation is specified more clearly than project funding vs. other evaluation situations.

What then, does “national” mean in national research evaluation systems? Research is evaluated at different levels and for different reasons. On the one hand, there is the national and international community of a given discipline that evaluates research from the disciplinary perspective, work by work. This is usually done by peer review for manuscript or presentation evaluation. It decides which research is accepted by the disciplinary community. On the other hand, there is the evaluation that is concerned with distribution of funding and procedural assessment at institutions. The second forms “national research evaluation system” and comprises several types of evaluation procedures as displayed in Table 1. The two perspectives on evaluation, the disciplinary and the more institutional one, are rarely distinguished clearly (see also Reymert et al., 2020). On the one hand, the scholars themselves adhere to the first and, when acting as experts, try to stick to this logic of evaluation, while the institution conducting the evaluation sets limits according to the function that the evaluation has to fulfil, which sometimes leads to tensions. At the same time, as scholars are regularly evaluated by institutions, they bring their experience of “national research evaluation” into the disciplinary evaluation (for example, with the introduction of societal impact in the REF, British reviewers suddenly put emphasis on societal relevance and potential impact in manuscript peer review of non-

British research; American reviewers might ask researchers to add “race” to their analysis, which might cause problems to European researchers, where the concept of “race” has a certain connotation). These intersections between national and international evaluation practices are not yet studied enough as the diversity of national evaluation practices and their impact on the perceptions of what scholars deem important has not received enough attention so far.

National Research Evaluation Systems and the Social Sciences

Until recently, research evaluation procedures have been usually implemented in top-down manner without paying attention to research practice and knowledge creation and much less so to disciplinary differences regarding the research process.

The methods applied in many research evaluation procedures are based – for historical reasons – on the knowledge creation process in some disciplines of the natural and life sciences and are especially inadequate for the use in the SSH disciplines (Hicks, 2004; Nederhof, 2006; Van Leeuwen, 2013; Hug et al., 2013). The problems are most pronounced for the humanities and therefore many studies focus on the humanities. But similar problems apply to the social sciences. Furthermore, as already mentioned, it is often not easy to differentiate clearly between social sciences and humanities. Bonaccorsi suggests in this book that epistemic tradition of the discipline should decide which evaluation procedure to apply. However, while taking epistemic differences in research practice into account in evaluation, it does not mean that applying evaluation methods from the natural sciences to other disciplines is adequate, simply because they try to mimic natural science methods. There is rather evidence for the fact that the current evaluation procedures lack validity in the natural sciences as well, as the negative effects become more evident (De Rijcke et al., 2016; DORA, 2012, <https://sfdora.org>).

National research evaluation initially mainly impacted the natural and life sciences due to their costly nature and close relationship with the industry. However, they are now also applied in the SSH (Guillory, 2005). The discontent of SSH scholars with the evaluation procedures ill-adapted to their disciplines led some scholars and policy makers to explore, in a bottom-up process, evaluation procedures that are better suited for SSH research (for overviews, see Ochsner et al., 2016, 2017). Such procedures rely more strongly on peer review, include a wider range of publication types and account for linguistic and cultural aspects of research (as is now also increasingly the recognised for all disciplines, see DORA case studies, 2020, <https://sfdora.org/dora-case-studies/>).

With regard to the relationship of national evaluation systems, the typologies have ignored the aspects of SSH with a few exceptions stating that in SSH, other publication databases are used (Zacharewicz et al., 2018) or that even though there are concerns about the adequacy of methods for the SSH, adaptations are rare (Von Tunzelmann & Kraemer Mbula, 2003). Only recently, this relationship has been investigated more thoroughly by two collaborative projects. Van Gestel and Lienhard (2019b) describe the evaluation practices in eleven countries with regard to legal science. They find that there is no transnational debate about what constitutes good legal research, methodology and how to evaluate it. Rather, it seems that across countries, legal research is often evaluated according to standards not suitable to the characteristics of legal scholarship. However, the authors do not suggest joining in the choir of complaint but rather see the lagging behind in evaluation practice as an advantage: legal scholars can learn from the mistakes made in other disciplines (Van Gestel & Lienhard, 2019a). The study of law crosses the boundaries between the social sciences and humanities, depending on one's view of research into its objectives and effects: This is why it is a challenging discipline to investigate on.

In their typology, Ochsner et al. (2018) explicitly include characteristics of SSH adaptations. Their classification includes whether there are SSH specific adaptations in institutional evaluation, whether there are specific calls for SSH topics in project funding, and whether there is a push to English publications. The results show that there is quite some variance with regard to how national research evaluation treats SSH disciplines. While only 8 out of 32 countries have a strong push to English

language publications (and all are Central and Eastern European countries), less than half of the countries allow for adaptations to SSH research practices (14 countries) or offer a dedicated project funding stream to SSH topics (13 countries). Such adaptations are not only the case in procedures mainly based on peer review. Countries that have a performance-based component in their evaluation procedures also allow for adaptations to SSH research practices, for example by including a wide range of publication types in their metrics. Book publications are especially important in SSH disciplines and their inclusion in evaluation procedures vary across countries (Giménez-Toledo et al., 2019).

Conclusions

In this chapter, we took a closer look at how research is evaluated across European countries. We have summarised current research on evaluation practices and grouped existing typologies into three groups. We found that the first typologies presented around the change of the millennium do not cover the important aspects of research evaluation in the respective countries by focusing on performance-based procedures only. We also concluded that a shift towards including project funding in the second group of typologies still provides a restricted comprehension on how research is evaluated across European countries. We suggest that the focus on a managerial perspective misses out an important fact when analysing research evaluation procedures: Evaluation aims at steering the behaviour of researchers in one way or the other. However, focusing on only one or two evaluation procedures neglects that researchers are acting in complex environments. The third group of typologies, therefore, takes the opposite perspective on research evaluation. Instead of taking the top-down managerial perspective, these studies take the bottom-up perspective from the shop floor and follow an analytical rather than a descriptive approach. They investigate how researchers and their practice are influenced by evaluations. This is highly relevant because if policy strives to incentivise some practices over others through evaluation, it must take research practice into account. Taking the bottom-up perspective, it becomes quickly evident that researchers are not influenced by one evaluation procedure but by a full set of different evaluation procedures. It is these complex *national evaluation systems* that need further scrutiny. This also reveals that, at the shop floor, contradictory incentives resulting from different assessment procedures might work on researchers simultaneously. On the other hand, incentives that come with risks of unintended effects on behaviour might be counter-balanced with other procedures incentivising behaviour that compensates the negative effects. Such a perspective on a “total effect of research evaluation procedures” helps to understand the ultimate goals of evaluation: to guide researchers in their activities. Thus, such a perspective focusing on the researchers adds value to policy makers and university managers as it reveals interdependencies of different evaluation procedures (and policies) that might seek to influence the same aspects in different ways either counterweighting or reinforcing each other.

Each country is confronted with its special situation and challenges. Research policy and evaluation must relate to this situation and the challenges. Therefore, the diversity must not be seen as a problem but as a functional way of organising and steering a specific research environment. Efforts to standardise research evaluation practices therefore could simplify the work of administrators and evaluators but will certainly not match the aim of evaluation: to provide researchers and university administrators with a tool to improve their organisation and research practice.

Practical implications

The analyses on how research is evaluated across countries relating to the SSH show that it is important to respect the diversity of research evaluation (see also Ochsner et al., 2020). This diversity has its roots in the diversity of research systems: the organisation of research is not the same in all countries, nor is the situation of the research landscape. Different evaluation procedures follow different goals and are linked to different values

Evaluation thus needs to reflect the policy needs of the country/institution. Evaluation has to correspond to policy goals as evaluation is only meaningful in its context of organisation, goals and

values (see also Dahler-Larsen, 2012, pp. 16-27). There is no “one and only true evaluation”; rather, evaluation is the negotiation of values linked to policy.

Evaluation needs to respect disciplinary differences in knowledge production. To pick up the seminal work of Dahler-Larsen (2012, p. 229), “evaluation machines” are not providing meaningful results as they are mechanistic, simplistic, forecast the past into the future and rarely can include the delicate nuances that are highly important in research. This is especially true when it concerns smaller disciplines, research relevant in a local context or dissemination practices that are not standardised, all of which are especially relevant for the social sciences.

While a standardisation of evaluation is thus not improving research evaluation, harmonisation might instead be increased. Harmonisation leaves the national differences in place and tries to bring them closer, respecting them. Harmonisation will help researchers to plan careers and to cooperate between countries. In such a context, scholars play an important role: As main actors of evaluation (in the double role of evaluated and evaluators) they respect research practices within countries, but they have also the right to work within a harmonised and coherent setting of principles, rules and policies which greatly influences the potential of future generations of researchers. Harmonisation includes the coordination of policy actions and measures in order to achieve goals such as greater efficiency and fairness which, especially in the context of career promotion, are very valuable.

Evaluation links different stakeholders. An important type of stakeholder is often excluded from the analysis: the evaluation industry itself (Ochsner et al., 2020). What is the role of service providers in research evaluation, such as Technopolis, Clarivate Analytics or Altmetrics.com? To what extent are their products that promise to produce comparable and easily interpretable results serving the goal of the evaluation exercise and to what extent are they merely cementing the companies’ position in the evaluation game? The mechanical aspects of such providers’ “evaluation machines” help simplify the work of the evaluators but deprive the (national) evaluation of its idiosyncrasies and the democratic process of negotiation of values (see Dahler-Larsen, 2012, pp. 236-237).

Evaluation can serve as a means to democratically negotiate values and norms about research and can help connect different stakeholders and contribute to the relevant discourses in a specific context. This, however, is only possible if the diversity of knowledge production, dissemination and evaluation practices is preserved.

Acknowledgements

This chapter is based upon work from COST Action CA 15137 ‘European Network for Research Evaluation in the SSH (ENRESSH)’, supported by COST (European Cooperation in Science and Technology).

References

- Aagaard, K. (2015). How incentives trickle down: Local use of a national bibliometric indicator system. *Science and Public Policy*, 42(5), 725-737. <http://doi.org/10.1093/scipol/scu087>
- Barré, R. (2010). Towards socially robust S&T indicators: Indicators as debatable devices, enabling collective learning. *Research Evaluation*, 19(3), 227-231. <http://doi.org/10.3152/095820210X512069>
- Bianco, M., Gras, N., & Sutz, J. (2016). Academic Evaluation: Universal Instrument? Tool for Development? *Minerva*, 54(4), 399-421. <http://doi.org/10.1007/s11024-016-9306-9>
- Burrows, R. (2012). Living with the h-index? Metric assemblages in the contemporary academy. *The Sociological Review*, 60(2), 355-372. <http://doi.org/10.1111/j.1467-954X.2012.02077.x>
- Clark, B. R. (1978). Academic Differentiation in National Systems of Higher Education. *Comparative Education Review*, 22(2), 242-258. <http://doi.org/10.2307/1187674>

- Coryn, C. L. S., Hattie, J. A., Scriven, M., & Hartmann, D. J. (2007). Models and Mechanisms for Evaluating Government-Funded Research: An International Comparison. *American Journal of Evaluation*, 28(4), 437-457. <http://doi.org/10.1177/1098214007308290>
- Dahler-Larsen, P. (2012). *The evaluation society*. Stanford University Press.
- De Rijcke, S., Holtrop, T., Kaltenbrunner, W., Zuijderwijk, J., Beaulieu, A., Franssen, T., Van Leeuwen, T., Monegon, P., Tatum, C., Valkenburg, G., & Wouters, P. (2019). Evaluative inquiry: Engaging research evaluation analytically and strategically. *Fteval Journal for Research and Technology Policy Evaluation*, 48, 176-182. <http://doi.org/10.22163/fteval.2019.386>
- De Rijcke, S., Wouters, P. F., Rushforth, A. D., Franssen, T. P., & Hammarfelt, B. (2016). Evaluation practices and effects of indicator use: A literature review. *Research Evaluation*, 25(2), 161-169. <http://doi.org/10.1093/reseval/rvv038>
- Deem, R., Hillyard, S., & Reed, M. (2007). *Knowledge, higher education, and the new managerialism: The changing management of UK universities*. Oxford University Press.
- European Commission. (2005). *Commission Recommendation of 11 March 2005 on the European Charter for Researchers and on a Code of Conduct for the Recruitment of Researchers (2005/251/EC)*. European Union. <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2005:075:0067:0077:EN:PDF>
- European Commission. (2011). *Towards a European Framework for Research Careers*. https://cdn5.euraxess.org/sites/default/files/policy_library/towards_a_european_framework_forresearch_careers_final.pdf
- European Commission. (2020). *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: A new ERA for Research and Innovation*. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0628&from=EN>
- Galleron, I., Ochsner, M., Spaapen, J., & Williams, G. (2017). Valorizing SSH research: Towards a new approach to evaluate SSH research' value for society. *Fteval Journal for Research and Technology Policy Evaluation*, 44, 35-41. <http://doi.org/10.22163/fteval.2017.274>
- Geuna, A., & Martin, B. R. (2001). *University Research Evaluation and Funding: An International Comparison*. SPRU Electronic Working Paper Series (Vol. 71). <http://www.sussex.ac.uk/Units/spru/publications/imprint/sewps/sewp71/sewp71.pdf>
- Geuna, A., & Martin, B. R. (2003). University Research Evaluation and Funding: An International Comparison. *Minerva*, 41(4), 277-304. <http://doi.org/10.1023/B:MINE.0000005155.70870.bd>
- Geuna, A., Hidayat, D., & Martin, B. R. (1999). *Resource Allocation and Research Performance: The Assessment of Research: A study carried out for the Higher Education Funding Council of England*. SPRU.
- Giménez-Toledo, E., Mañana-Rodríguez, J., Engels, T. C. E., Guns, R., Kulczycki, E., Ochsner, M., Pölönen, J., Sivertsen, G., & Zuccala, A. (2019). Taking scholarly books into account, part II: A comparison of 19 European countries in evaluation and funding. *Scientometrics*, 118(1), 233-251. <http://doi.org/10.1007/s11192-018-2956-7>

- Guillory, J. (2005). Valuing the humanities, evaluating scholarship. *Profession, 11*, 28-38. <https://www.jstor.org/stable/25595795>
- Hicks, D. (2004). The four literatures of social science. In H. F. Moed, W. Glanzel, & U. Schmoch (Eds.), *Handbook of quantitative science and technology research* (pp. 473-496). Kluwer Academic.
- Hicks, D. (2010). Overview of models of performance-based research funding systems. In *Performance-based Funding for Public Research in Tertiary Education Institutions* (pp. 23-52). OECD. <http://doi.org/10.1787/9789264094611-4-en>
- Hicks, D. (2012). Performance-based university research funding systems. *Research Policy, 41*(2), 251-261. <http://doi.org/10.1016/j.respol.2011.09.007>
- Hug, S. E., Ochsner, M., & Daniel, H. D. (2013). Criteria for assessing research quality in the humanities: A Delphi study among scholars of English literature, German literature and art history. *Research Evaluation, 22*(5), 369-383. <http://doi.org/10.1093/reseval/rvt008>
- Jonkers, K., & Zacharewicz, T. (2016). *Research Performance Based Funding Systems: A Comparative Assessment*. JRC101043. European Union. <https://dx.doi.org/10.2760/70120>
- Lariviere, V., Gingras, Y., & Archambault, É. (2006). Canadian collaboration networks: A comparative analysis of the natural sciences, social sciences and the humanities. *Scientometrics, 68*(3), 519-533. <http://doi.org/10.1007/s11192-006-0127-8>
- Lepori, B., Reale, E., & Spinello, A. O. (2018). Conceptualizing and measuring performance orientation of research funding systems. *Research Evaluation, 27*(3), 171-183. <http://doi.org/10.1093/reseval/rvy007>
- Lepori, B., Van den Besselaar, P., Dinges, M., Poti, B., Reale, E., Slipersæter, S., Thèves, J., & Van der Meulen, B. (2007). Comparing the evolution of national research policies: What patterns of change? *Science and Public Policy, 34*(6), 372-388. <http://doi.org/10.3152/030234207X234578>
- Lock, G., & Martins, H. (2011). Quantified Control and the Mass Production of "Psychotic Citizens". *Revue Électronique Des Sciences Humaines Et Sociales*. <https://www.espacestemps.net/articles/quantified-control-and-the-mass-production-of-psychotic-citizensrdquo/>
- Nederhof, A. J. (2006). Bibliometric monitoring of research performance in the social sciences and the humanities: A review. *Scientometrics, 66*(1), 81-100. <http://doi.org/10.1007/s11192-006-0007-2>
- Ochsner, M., Hug, S. E., & Daniel, H. D. (2016). *Research Assessment in the Humanities: Towards Criteria and Procedures*. Springer. <http://doi.org/10.1007/978-3-319-29016-4>
- Ochsner, M., Hug, S. E., & Galleron, I. (2017). The future of research assessment in the humanities: Bottom-up assessment procedures. *Palgrave Communications, 3*, Article 17020. <https://doi.org/10.1057/PALCOMMS.2017.20>
- Ochsner, M., Kulczycki, E., & Gedutis, A. (2018). The Diversity of European Research Evaluation Systems. In R. Costas, T. Franssen, & A. Yegros-Yegros (Eds.), *Proceedings of the 23rd International Conference on Science and Technology Indicators* (pp. 1234-1241). CWTS.

- Ochsner, M., Kulczycki, E., Gedutis, A., & Peruginelli, G. (2021). National Research Evaluation Systems. In R. Ball (Ed.), *Handbook Bibliometrics* (pp. 99-106). De Gruyter Saur. <http://doi.org/10.1515/9783110646610-01>
- Ochsner, M., Ma, L., Kancewicz-Hoffman, N., Holm, J., Gedutis, A., Šima, K., Hug, S. E., Dewaele, A., & De Jong, S. (2020). *ENRESSH Policy Brief Research Evaluation: Better adapted procedures for research evaluation in the SSH*. ENRESSH. <https://doi.org/10.6084/m9.figshare.12049314.v1>
- Plumpe, W. (2009). *Qualitätsmessung: Stellungnahme zum Rating des Wissenschaftsrates aus Sicht des Historikerverbandes*. H-SOZ-KULT. <https://www.hsozkult.de/debate/id/diskussionen-1101>
- Reale, E. (2017). *Analysis of national public research funding (PREF). Final report*. JRC107599. Publications Office of the European Union. <http://dx.doi.org/10.2760/19140>
- Reymert, I., Jungblut, J., & Borlaug, N. S. B. (2020). Are evaluative cultures national or global? A cross-national study on evaluative cultures in academic recruitment processes in Europe. *Higher Education*. Advance online publication. <http://doi.org/10.1007/s10734-020-00659-3>
- Rolfe, G. (2013). *The university in dissent: Scholarship in the corporate university*. Routledge.
- Saenen, B., Morais, R., Gaillard, V., & Borrell-Damián, L. (2019). *Research Assessment in the Transition to Open Science: 2019 EUA Open Science and Access Survey Results*. EUA. <https://eua.eu/resources/publications/888:research-assessment-in-the-transition-to-open-science.html>.
- Science Europe. (2020). *Position Statement and Recommendations on Research Assessment Processes*. Science Europe. <https://www.scienceurope.org/our-resources/position-statement-research-assessment-processes/>.
- Sievers, B. (2008). The Psychotic University. *Ephemer*, 8(3), 238-257. <http://www.ephemerajournal.org/contribution/psychotic-university>
- Snel, M. (2019). Evaluation of academic legal publications at the EU level. In R. Van Gestel & A. Lienhard (Eds.), *Evaluating academic legal research in Europe: The advantage of lagging behind* (pp. 384-421). Edward Elgar.
- Torrance, H. (2020). The Research Excellence Framework in the United Kingdom: Processes, Consequences, and Incentives to Engage. *Qualitative Inquiry*, 26(7), 771-779. <http://doi.org/10.1177/1077800419878748>
- Van Gestel, R., & Lienhard, A. (Eds.). (2019a). Conclusion and Discussion. In R. Van Gestel & A. Lienhard (Eds.), *Evaluating academic legal research in Europe: The advantage of lagging behind* (422-460). Edward Elgar.
- Van Gestel, R., & Lienhard, A. (Eds.). (2019b). *Evaluating academic legal research in Europe: The advantage of lagging behind*. Edward Elgar.
- Van Leeuwen, T. N. (2013). Bibliometric research evaluations, Web of Science and the Social Sciences and Humanities: A problematic relationship? *Bibliometrie - Praxis Und Forschung*, 2(8), 1-18. <https://doi.org/10.5283/bpf.173>

Von Tunzelmann, N., & Kraemer Mbula, E. (2003). *Changes in research assessment practices in other countries since 1999: Final report*. Bristol. <http://www.ra-review.ac.uk/reports/prac/changingpractices.pdf>

Whitley, R. (2003). Competition and pluralism in the public sciences: The impact of institutional frameworks on the organisation of academic science. *Research Policy*, 32(6), 1015-1029. [http://doi.org/10.1016/S0048-7333\(02\)00112-9](http://doi.org/10.1016/S0048-7333(02)00112-9)

Whitley, R. (2007). Changing governance of the public sciences. In R. Whitley & J. Gläser (Eds.), *The changing governance of the sciences: The advent of research evaluation systems* (pp. 3-27). Springer.

Zacharewicz, T., Lepori, B., Reale, E., & Jonkers, K. (2019). Performance-based research funding in EU Member States—a comparative assessment. *Science and Public Policy*, 46(1), 105-115.

DRAFT